**RESULTS BASED MANAGEMENT
IN THE DEVELOPMENT CO-OPERATION AGENCIES:
A REVIEW OF EXPERIENCE**

BACKGROUND REPORT

In order to respond to the need for an overview of the rapid evolution of RBM, the DAC Working Party on Aid Evaluation initiated a study of performance management systems. The ensuing draft report was presented to the February 2000 meeting of the WP-EV and the document was subsequently revised.

It was written by Ms. Annette Binnendijk, consultant to the DAC WP-EV.

This review constitutes the first phase of the project; a second phase involving key informant interviews in a number of agencies is due for completion by November 2001.

TABLE OF CONTENTS

PREFACE

At the meeting of the DAC Working Party on Aid Evaluation (WP-EV) held in January 1999, Members agreed to several follow-up activities to the *Review of the DAC Principles for Evaluation of Development Assistance*. One of the new areas of work identified was performance management systems. The DAC Secretariat agreed to lead and co-ordinate the work.

The topic of performance management, or results based management, was selected because many development co-operation agencies are now in the process of introducing or reforming their performance management systems and measurement approaches, and face a number of common issues and challenges. For example, how to establish an effective performance measurement system, deal with analytical issues of attributing impacts and aggregating results, ensure a distinct yet complementary role for evaluation, and establish organizational incentives and processes that will stimulate the use of performance information in management decision-making.

The objective of the work on performance management is "to provide guidance, based on Members' experience, on how to develop and implement results based management in development agencies and make it best interact with evaluation systems."[1]

This work on performance management is to be implemented in two phases:

- A review of the initial experiences of the development co-operation agencies with performance management systems.

- The development of "good practices" for establishing effective performance management systems in these agencies.

This paper is the product of the first phase. It is based on a document review of the experiences and practices of selected Member development co-operation agencies with establishing performance or results based management systems. The paper draws heavily on discussions and papers presented at the Working Party's October 1998 *Workshop on Performance Management and Evaluation* sponsored by Sida and UNDP, and also on other recent documents updating performance management experiences and practices obtained from selected Members during the summer of 1999. (See annex for list of references).

A draft of this paper was submitted to Members of the DAC Working Party on Aid Evaluation in November 1999 and was reviewed at the February 2000 meeting in Paris. Members' comments from that meeting have been incorporated into this revised version, dated October 2000.

The development co-operation (or donor) agencies whose experiences are reviewed include USAID, DFID, AusAID, CIDA, Danida, the UNDP and the World Bank. These seven agencies made presentations on their performance management systems at the October 1998 workshop and have considerable documentation concerning their experiences. (During the second phase of work, the relevant experiences of other donor agencies will also be taken into consideration).

1.　　See *Complementing and Reinforcing the DAC Principles for Aid Evaluation* [DCD/DAC/EV(99)5], p. 6.

This paper synthesizes the experiences of these seven donor agencies with establishing and implementing their results based management systems, comparing similarities and contrasting differences in approach. Illustrations drawn from individual donor approaches are used throughout the paper. Key features of results based management are addressed, beginning with the phases of performance measurement -- e.g., clarifying objectives and strategies, selecting indicators and targets for measuring progress, collecting data, and analyzing and reporting results achieved. Performance measurement systems are examined at three key organizational levels -- the traditional project level, the country program level, and the agency-wide (corporate or global) level. Next, the role of evaluation vis-à-vis performance measurement is addressed. Then the paper examines how the donor agencies use performance information -- for external reporting, and for internal management learning and decision-making processes. It also reviews some of the organizational mechanisms, processes and incentives used to help ensure effective use of performance information, e.g., devolution of authority and accountability, participation of stakeholders and partners, focus on beneficiary needs and preferences, creation of a learning culture, etc. The final section outlines some conclusions and remaining challenges, offers preliminary lessons, and reviews next steps being taken by the Working Party on Aid Evaluation to elaborate good practices for results based management in development co-operation agencies.

Some of the key topics discussed in this paper include:

- Using analytical frameworks for formulating objectives and for structuring performance measurement systems.

- Developing performance indicators -- types of measures, selection criteria, etc.

- Using targets and benchmarks for judging performance.

- Balancing the respective roles of implementation and results monitoring.

- Collecting data -- methods, responsibilities, harmonization, and capacity building issues.

- Aggregating performance (results) to the agency level.

- Attributing outcomes and impacts to a specific project, program, or agency.

- Integrating evaluation within the broader performance management system.

- Using performance information -- for external performance reporting to stakeholders and for internal management learning and decision-making processes.

- Stimulating demand for performance information via various organizational reforms, mechanisms, and incentives.

I. RESULTS BASED MANAGEMENT IN OECD COUNTRIES

-- An Overview of Key Concepts, Definitions and Issues --

Public sector reforms

During the 1990s, many of the OECD countries have undertaken extensive public sector reforms in response to economic, social and political pressures. For example, common economic pressures have included budget deficits, structural problems, growing competitiveness and globalization. Political and social factors have included a lack of public confidence in government, growing demands for better and more responsive services, and better accountability for achieving results with taxpayers' money. Popular catch phrases such as "Reinventing government", "Doing more with less", "Demonstrating value for money", etc. describe the movement towards public sector reforms that have become prevalent in many of the OECD countries.

Often, government-wide legislation or executive orders have driven and guided the public sector reforms. For example, the passage of the 1993 *Government Performance and Results Act* was the major driver of federal government reform in the United States. In the United Kingdom, the publication of a 1995 White Paper on *Better Accounting for the Taxpayers' Money* was a key milestone committing the government to the introduction of resource accounting and budgeting. In Australia the main driver for change was the introduction of Accruals-based Outcome and Output Budgeting. In Canada, the Office of the Auditor General and the Treasury Board Secretariat have been the primary promoters of reforms across the federal government.

While there have been variations in the reform packages implemented in the OECD countries, there are also many common aspects found in most countries, for example:

- Focus on performance issues (e.g. efficiency, effectiveness, quality of services).
- Devolution of management authority and responsibility.
- Orientation to customer needs and preferences.
- Participation by stakeholders.
- Reform of budget processes and financial management systems.
- Application of modern management practices.

Results based management (performance management)

Perhaps the most central feature of the reforms has been the emphasis on improving performance and ensuring that government activities achieve desired results. A recent study of the experiences of ten OECD Member countries with introducing performance management showed that it was a key feature in the reform efforts of all ten. [2]

Performance management, also referred to as results based management, can be defined as a broad management strategy aimed at achieving important changes in the way government agencies operate, with improving performance (achieving better results) as the central orientation.

Performance measurement is concerned more narrowly with the production *or supply* of performance information, and is focused on technical aspects of clarifying objectives, developing indicators, collecting and analyzing data on results. Performance management encompasses performance measurement, but is broader. It is equally concerned with generating management *demand* for performance information -- that is, with its uses in program, policy, and budget decision-making processes and with establishing organizational procedures, mechanisms and incentives that actively encourage its use. In an effective performance management system, achieving results and continuous improvement based on performance information is central to the management process.

Performance measurement

Performance measurement is the process an organization follows to objectively measure how well its stated objectives are being met. It typically involves several phases: e.g., articulating and agreeing on objectives, selecting indicators and setting targets, monitoring performance (collecting data on results), and analyzing those results vis-à-vis targets. In practice, results are often measured without clear definition of objectives or detailed targets. As performance measurement systems mature, greater attention is placed on measuring what's important rather than what's easily measured. Governments that emphasize accountability tend to use performance targets, but too much emphasis on "hard" targets can potentially have dysfunctional consequences. Governments that focus more on management improvement may place less emphasis on setting and achieving targets, but instead require organizations to demonstrate steady improvements in performance/ results.

Uses of performance information

The introduction of performance management appears to have been driven by two key aims or intended uses -- management improvement and performance reporting (accountability). In the first, the focus is on using performance information for management learning and decision-making processes. For example, when managers routinely make adjustments to improve their programs based on feedback about results being achieved. A special type of management decision-making process that performance information is increasingly being used for is resource allocation. In performance based budgeting, funds are allocated across an agency's programs on the basis of results, rather than inputs or activities. In the second aim, emphasis shifts to holding managers accountable for achievement of specific planned results or targets, and to transparent reporting of

---

2.　　　See *In Search of Results: Public Management Practices* (OECD, 1997).

those results. In practice, governments tend to favor or prioritize one or the other of these objectives. To some extent, these aims may be conflicting and entail somewhat different management approaches and systems.

When performance information is used for reporting to external stakeholder audiences, this is sometimes referred to as *accountability-for-results*. Government-wide legislation or executive orders often mandate such reporting. Moreover, such reporting can be useful in the competition for funds by convincing a sceptical public or legislature that an agency's programs produce significant results and provide "value for money". Annual performance reports may be directed to many stakeholders, for example, to ministers, parliament, auditors or other oversight agencies, customers, and the general public.

When performance information is used in internal management processes with the aim of improving performance and achieving better results, this is often referred to as *managing-for-results*. Such actual use of performance information has often been a weakness of performance management in the OECD countries. Too often, government agencies have emphasized performance measurement for external reporting only, with little attention given to putting the performance information to use in internal management decision-making processes.

For performance information to be used for management decision-making requires that it becomes integrated into key management systems and processes of the organization; such as in strategic planning, policy formulation, program or project management, financial and budget management, and human resource management.

Of particular interest is the intended use of performance information in the budget process for improving budgetary decisions and allocation of resources. The ultimate objective is ensuring that resources are allocated to those programs that achieve the best results at least cost, and away from poor performing activities. Initially, a more modest aim may be simply to estimate the costs of achieving planned results, rather than the cost of inputs or activities, which has been the traditional approach to budgeting. In some OECD countries, performance-based budgeting is a key objective of performance management. However, it is not a simple or straightforward process that can be rigidly applied. While it may appear to make sense to reward organizations and programs that perform best, punishing weaker performers may not always be feasible or desirable. Other factors besides performance, especially political considerations, will continue to play a role in budget allocations. However, performance measurement can become an important source of information that feeds into the budget decision-making process, as one of several key factors.

However, these various uses of performance information may not be completely compatible with one another, or may require different types or levels of result data to satisfy their different needs and interests. Balancing these different needs and uses without over-burdening the performance management system remains a challenge.

Role of evaluation in performance management

The role of evaluation vis-à-vis performance management has not always been clear-cut. In part, this is because evaluation was well established in many governments before the introduction of performance management and the new approaches did not necessarily incorporate evaluation. New performance management techniques were developed partly in response to perceived failures of evaluation; for example, the perception that uses of evaluation findings were limited relative to their costs. Moreover, evaluation was often viewed as a specialized function carried out by external experts or independent units, whereas performance

management, which involves reforming core management processes, was essentially the responsibility of managers within the organization.

Failure to clarify the relationship of evaluation to performance management can lead to duplication of efforts, confusion, and tensions among organizational units and professional groups. For example, some evaluators are increasingly concerned that emphasis on performance measurement may be replacing or "crowding out" evaluation in U.S. federal government agencies.

Most OECD governments see evaluation as part of the overall performance management framework, but the degree of integration and independence varies. Several approaches are possible.

At one extreme, evaluation may be viewed as a completely separate and independent function with clear roles vis-à-vis performance management. From this perspective, performance management is like any other internal management process that has to be subjected to independent evaluation. At the other extreme, evaluation is seen not as a separate or independent function but as completely integrated into individual performance management instruments.

A middle approach views evaluation as a separate or specialized function, but integrated into performance management. Less emphasis is placed on independence, and evaluation is seen as one of many instruments used in the overall performance management framework. Evaluation is viewed as complementary to -- and in some respects superior to -- other routine performance measurement techniques. For example, evaluation allows for more in-depth study of program performance, can analyze causes and effects in detail, can offer recommendations, or may assess performance issues normally too difficult, expensive or long-term to assess through on-going monitoring.

This middle approach has been gaining momentum. This is reflected in PUMA's *Best Practice Guidelines for Evaluation* (OECD, 1998) which was endorsed by the Public Management Committee. The Guidelines state that "evaluations must be part of a wider performance management framework". Still, some degree of independent evaluation capacity is being preserved; such as most evaluations conducted by central evaluation offices or performance audits carried out by audit offices. There is also growing awareness about the benefits of incorporating evaluative methods into key management processes. However, most governments see this as supplementing, rather than replacing more specialized evaluations.

II. RESULTS BASED MANAGEMENT
IN THE DEVELOPMENT CO-OPERATION AGENCIES

Introduction

As has been the case more broadly for the public sector of the OECD countries, the development co-operation (or donor) agencies have faced considerable external pressures to reform their management systems to become more effective and results-oriented. "Aid fatigue", the public's perception that aid programs are failing to produce significant development results, declining aid budgets, and government-wide reforms have all contributed to these agencies' recent efforts to establish results based management systems.

Thus far, the donor agencies have gained most experience with establishing performance measurement systems -- that is, with the provision of performance information -- and some experience with external reporting on results. Experience with the actual use of performance information for management decision-making, and with installing new organizational incentives, procedures, and mechanisms that would promote its internal use by managers, remains relatively weak in most cases.

Features and phases of results based management

Donor agencies broadly agree on the definition, purposes, and key features of results based management systems. Most would agree, for example, with quotes such as these:

- "Results based management provides a coherent framework for strategic planning and management based on learning and accountability in a decentralised environment. It is first a management system and second, a performance reporting system."[3]

- "Introducing a results-oriented approach ... aims at improving management effectiveness and accountability by defining realistic expected results, monitoring progress toward the achievement of expected results, integrating lessons learned into management decisions and reporting on performance."[4]

---

3.   *Note on Results Based Management*, Operations Evaluation Department, World Bank, 1997.

4.   *Results Based Management in Canadian International Development Agency*, CIDA, January 1999.

The basic purposes of results based management systems in the donor agencies are to generate and use performance information for accountability reporting to external stakeholder audiences and for internal management learning and decision-making. Most agencies' results based management systems include the following processes or phases:[5]

1. ***Formulating objectives:*** Identifying in clear, measurable terms the results being sought and developing a conceptual framework for how the results will be achieved.

2. ***Identifying indicators:*** For each objective, specifying exactly what is to be measured along a scale or dimension.

3. ***Setting targets:*** For each indicator, specifying the expected or planned levels of result to be achieved by specific dates, which will be used to judge performance.

4. ***Monitoring results:*** Developing performance monitoring systems to regularly collect data on actual results achieved.

5. ***Reviewing and reporting results***: Comparing actual results vis-à-vis the targets (or other criteria for making judgements about performance).

6. ***Integrating evaluations:*** Conducting evaluations to provide complementary information on performance not readily available from performance monitoring systems.

7. ***Using performance information:*** Using information from performance monitoring and evaluation sources for internal management learning and decision-making, and for external reporting to stakeholders on results achieved. Effective use generally depends upon putting in place various organizational reforms, new policies and procedures, and other mechanisms or incentives.

The first three phases or processes generally relate to a results-oriented planning approach, sometimes referred to as *strategic planning*. The first five together are usually included in the concept of *performance measurement.* All seven phases combined are essential to an effective *results based management system.* That is, integrating complementary information from both evaluation and performance measurement systems and ensuring management's use of this information are viewed as critical aspects of results based management. (See Box 1.)

Other components of results based management

In addition, other significant reforms often associated with results based management systems in development co-operation agencies include the following. Many of these changes in act to stimulate or facilitate the use of performance information.

•   *Holding managers accountable*: Instituting new mechanisms for holding agency managers and staff accountable for achieving results within their sphere of control.
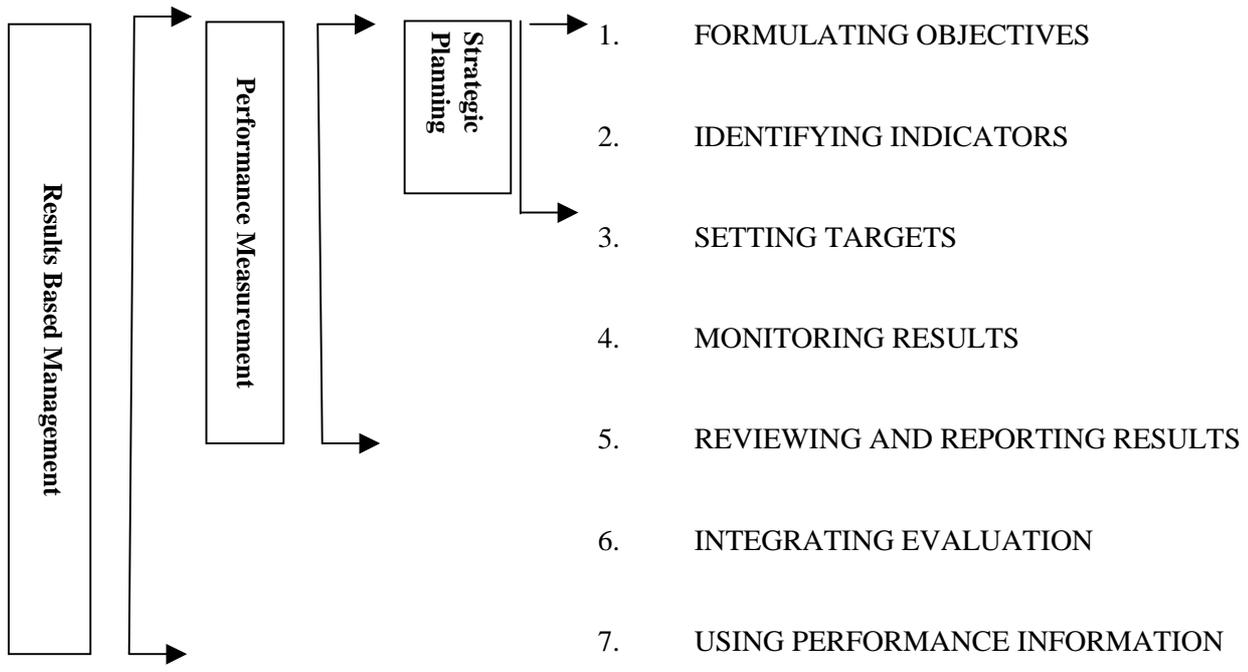
---

5.   These phases are largely sequential processes, but may to some extent proceed simultaneously.

- *Empowering managers*: Delegating authority to the management level being held accountable for results – thus empowering them with flexibility to make corrective adjustments and to shift resources from poorer to better performing activities.

- *Focusing on clients*: Consulting with and being responsive to project/program beneficiaries or clients concerning their preferences and satisfaction with goods and services provided.

- *Participation and partnership:* Including partners (e.g., from implementing agencies, partner country organizations, other donor agencies) that have a shared interest in achieving a development objective in all aspects of performance measurement and management processes. Facilitating putting partners from developing countries "in the driver's seat", for example by building capacity for performance monitoring and evaluation.

- *Reforming policy and procedure*: Officially instituting changes in the way the donor agency conducts its business operations by issuing new policies and procedural guidelines on results based management. Clarifying new operational procedures, roles and responsibilities.

- *Developing supportive mechanisms*: Assisting managers to effectively implement performance measurement and management processes, by providing appropriate training and technical assistance, establishing new performance information databases, developing guidebooks and best practices series.

- *Changing organizational culture*: Facilitating changes in the agency's culture – i.e., the values, attitudes, and behaviors of its personnel - required for effectively implementing results based management. For example, instilling a commitment to honest and open performance reporting, re-orientation away from inputs and processes towards results achievement, encouraging a learning culture grounded in evaluation, etc.

Results based management at different organizational levels

Performance measurement, and results based management more generally, takes place at different organizational or management levels within the donor agencies. The first level, which has been established the longest and for which there is most experience, is at the project level. More recently, efforts have been underway in some of the donor agencies to establish country program level performance measurement and management systems within their country offices or operating units. Moreover, establishing performance measurement and management systems at the third level -- the corporate or agency-wide level -- is now taking on urgency in many donor agencies as they face increasing public pressures and new government-wide legislation or directives to report on agency performance.

Box 1: Seven Phases of Results Based Management

Results Based Management	**Performance Measurement**	**Strategic Planning**	1. FORMULATING OBJECTIVES
			2. IDENTIFYING INDICATORS
			3. SETTING TARGETS
			4. MONITORING RESULTS
			5. REVIEWING AND REPORTING RESULTS
			6. INTEGRATING EVALUATION
			7. USING PERFORMANCE INFORMATION

Box 2 illustrates the key organizational levels at which performance measurement and management systems may take place within a donor agency.

**Box 2: Results Based Management
at Different Organizational Levels**

Agency-Wide
Level

Country Program
Level

Project Level
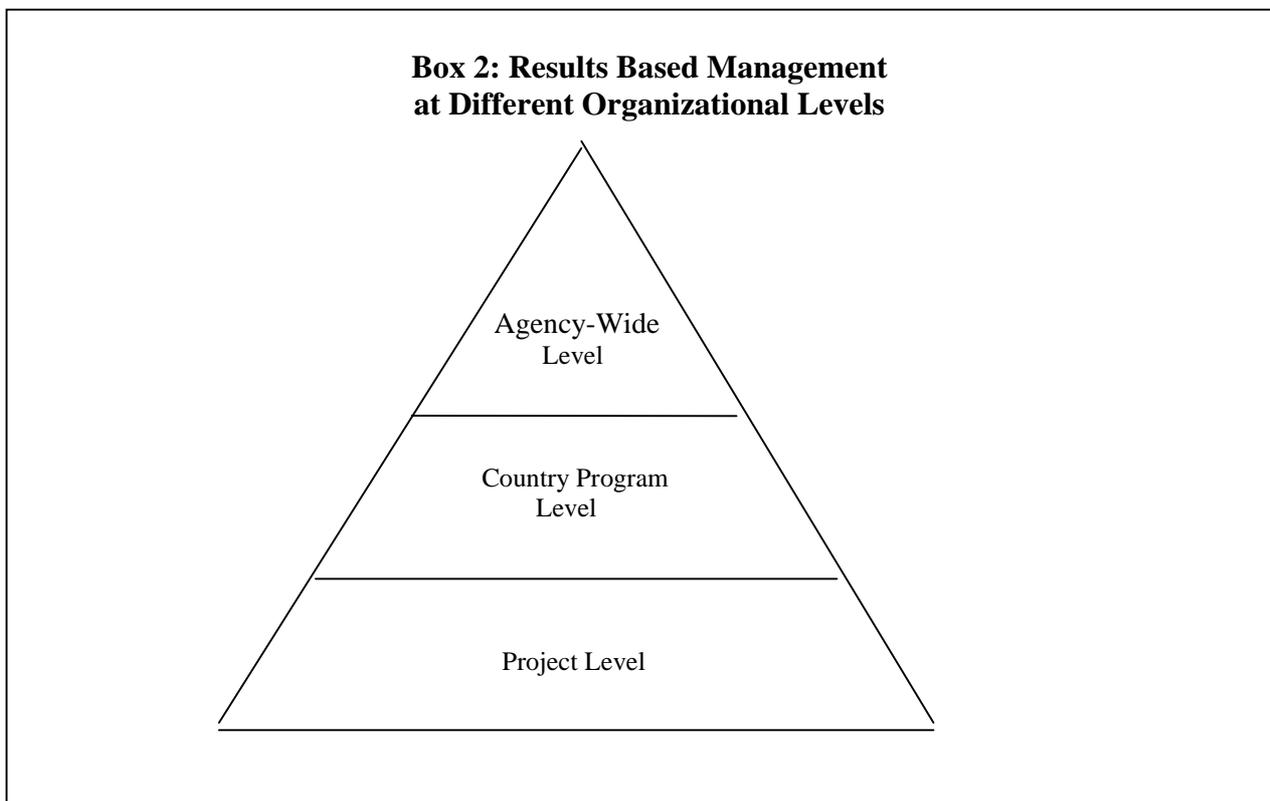
Donor agencies reviewed

The donor agencies reviewed in this paper were selected because they had considerable experience with (and documentation about) establishing a results based management system. They include five bilateral and two multilateral agencies:

- USAID (United States)
- DFID (United Kingdom)
- AusAID (Australia)
- CIDA (Canada)
- Danida (Denmark)
- UNDP
- World Bank

Certainly other donor agencies may also have relevant experiences, perhaps just not "labeled" as results based management. Still others may be in the beginning stages of introducing results based management systems but do not yet have much documentation about their early experiences. Additional agencies' experiences will be covered in the second phase of work on results based management.

Special challenges facing the donor agencies

Because of the nature of development co-operation work, the donor agencies face special challenges in establishing their performance management and measurement systems. These challenges are in some respects different from, and perhaps more difficult than, those confronting most other domestic government agencies.[6] This can make establishing performance measurement systems in donor agencies more complex and costly than normal. For example, donor agencies:

- Work in many different countries and contexts.

- Have a wide diversity of projects in multiple sectors.

- Often focus on capacity building and policy reform, which are harder to measure than direct service delivery activities.

- Are moving into new areas such as good governance, where there is little performance measurement experience.

- Often lack standard indicators on results/outcomes that can be easily compared and aggregated across projects and programs.

- Are usually only one among many partners contributing to development objectives, with consequent problems in attributing impacts to their own agency's projects and programs.

- Typically rely on results data collected by partner countries, which have limited technical capacity with consequent quality, coverage and timeliness problems.

- Face a greater potential conflict between the performance information demands of their own domestic stakeholders (e.g., donor country legislators, auditors, tax payers) versus the needs, interests and capacities of their developing country partners.

In particular, a number of these factors can complicate the donor agencies' efforts to compare and aggregate results across projects and programs to higher organizational and agency-wide levels.

Organization of the paper

The next three chapters focus on the experiences of the selected donor agencies with establishing their performance measurement systems, at the project, country program, and agency-wide levels. The subsequent chapter deals with developing a complementary role for evaluation vis-à-vis the performance measurement system. Next, there is a chapter examining issues related to the demand for performance information (from performance monitoring and evaluation sources) -- such as (a) the types of uses to which it is put and (b) the organizational policies and procedures, mechanisms, and incentives that can be established to encourage its use. The final chapter highlights some conclusions and remaining challenges, offers preliminary lessons about effective practices, and discusses the DAC Working Party on Aid Evaluation's next phase of work on results based management systems.

---

6.   Of course, it is not at all easy to conduct performance measurement for some other government functions, such as defence, foreign affairs, basic scientific research, etc.

III. PERFORMANCE MEASUREMENT IN THE DEVELOPMENT CO-OPERATION AGENCIES

-- The Project Level --

Many of the development co-operation agencies are now either designing, installing or reforming their performance measurement systems. Others are considering such systems. Thus, they are struggling with common problems of how to institute effective processes and practices for measuring their performance.

All seven of the donor agencies reviewed have had considerable experience with performance measurement at the project level. Well-established frameworks, systems and practices have, for the most part, been in place for some years. There is a good deal of similarity in approach among agencies at the project level. Most agencies have also initiated performance measurement systems at higher or more comprehensive organizational levels as well -- such as at the country program level and/or at the agency-wide (corporate) level. But, generally speaking, experience at these levels is more recent and less well advanced. Yet, establishing measurement systems at these higher organizational levels -- particularly at the corporate level -- is currently considered an urgent priority in all the agencies reviewed. Agency level performance measurement systems are necessary to respond to external domestic pressures to demonstrate the effectiveness in achieving results of the development assistance program as a whole. How to effectively and convincingly link performance across these various levels via appropriate aggregation techniques is currently a major issue and challenge for these agencies.

This chapter focuses on the development agencies' approach to performance measurement at the project level – where there is the most experience. Subsequent chapters review initial efforts at the country program and corporate levels.

Performance measurement at the project level

Performance measurement at the project level is concerned with measuring both a project's implementation progress and with results achieved. These two broad types of project performance measurement might be distinguished as (1) *implementation measurement* which is concerned with whether project inputs (financial, human and material resources) and activities (tasks, processes) are in compliance with design budgets, workplans, and schedules, and (2) *results measurement* which focuses on the achievement of project objectives (i.e., whether actual results are achieved as planned or targeted). Results are usually measured at three levels -- immediate outputs, intermediate outcomes and long-term impacts.[7] Whereas traditionally the development agencies focused mostly on implementation concerns, as they embrace results based management their focus is increasingly on measurement of results. Moreover, emphasis is shifting from immediate results (outputs) to medium and long-term results (outcomes, impacts).

---

7.   Some donor agencies (e.g., CIDA, USAID) use the term performance monitoring only in reference to the monitoring of results, not implementation. However, in this paper performance measurement and monitoring refers broadly to both implementation and results monitoring, since both address performance issues, although different aspects.

Overview of phases of performance measurement at the project level

Measuring performance at the project level can be divided into five processes or phases, as briefly outlined below:

1. **Formulating objectives:** As part of project planning, the project's objectives should be clarified by defining precise and measurable statements concerning the results to be achieved (outputs, purpose, and goal) and then identifying the strategies or means (inputs and activities) for meeting those objectives. The project logical framework, or logframe for short, is a favourite tool used by development agencies for conceptualizing a project's objectives and strategies. The logframe is typically based on a five-level hierarchy model with assumed cause-effect relationships among them, with those at the lower level of the hierarchy contributing to the attainment of those above. The logic is as follows: *inputs* are used to undertake project *activities* that lead to the delivery of *outputs (goods/services)*, that lead to the attainment of the *project purpose* that contributes to a *project goal*.

2. **Selecting indicators:** Next, indicators are developed for measuring implementation progress and achievement of results. The logframe provides a five-level structure around which the indicators are typically constructed. Indicators specify what to measure along a scale or dimension (e.g., numbers of workshops held, percent of farmers adopting new technology, ratio of female to male students, etc.). The relative importance of indicator types is likely to change over the project's life cycle, with more emphasis given at first to input and process indicators, while shifting later to output, outcome (purpose-level), and impact (goal-level) indicators.

3. **Setting targets**: Once indicators have been identified, actual baseline values should be collected for each, ideally just before the project gets underway. This will be important for gauging whether progress is being made later. Often agencies also set explicit targets for their indicators. A target specifies a particular value for an indicator to be accomplished within a given time frame. (For example, child immunization rates increased to 80 percent of children by 2003.). Targets help clarify exactly what needs to be accomplished by when. It represents a commitment and can help orient and motivate project staff and mangers to the tasks at hand.

4. **Monitoring (collecting) performance data**: Once indicators and targets are set, actual data for each indicator is collected at regular intervals. Implementation monitoring involves the on-going recording of data on project operations -- e.g., tracking funds and other inputs, and processes. It involves keeping good financial accounts and field activity records, and frequent checks to assess compliance with workplans and budgets. Results monitoring involves the periodic collection of data on the project's actual achievement of results -- e.g. its short-term outputs, medium-term outcomes, and long-term impacts. Data on project outputs are generated mostly by project staff and are based on simple reporting systems. Data on intermediate outcomes are generally collected from low-cost rapid appraisal methods, mini-surveys or consultations with project clients. Measuring impacts usually require conducting expensive sample surveys or relying on already existing data sources such as national surveys, censuses, registration systems, etc. Data collection at the higher levels -- especially at the impact level -- is often considered beyond the scope of the implementing agency's normal responsibility. Donor agencies will need to make special arrangements with partner country statistical organizations with data collection expertise for conducting or adding-on to planned surveys. Since several donor agencies working in the same sector may share needs for similar impact-level data, it would be useful to consider co-ordinating or jointly supporting these data collection efforts, to avoid duplication of effort and to share costs. Moreover, to ensure valid and reliable data, supporting capacity-building efforts may be called for as well.

5. **Reviewing and reporting performance data**: Review of project performance monitoring data most typically involves simple analysis comparing actual results achieved against planned results or targets. Not all agencies use targets, however. Some may look instead for continuous improvements and positive movement towards objectives, or make comparisons with similar projects known for their good performance. Using targets tends to imply management accountability for achieving them. While targets may be appropriate for outputs, and perhaps even for intermediate outcomes, their appropriateness for the goal/impact level might be questioned, given project management's very limited sphere of control or influence at this level. Analysis of performance monitoring data may address a broad variety of issues. Periodic reviews of performance data by project management will help alert them to problems, which may lead directly to taking actions or signal the need for more in-depth evaluation studies focused on specific performance issues.

The donor agencies' policies emphasize the importance of encouraging participation from the project implementing agency, the partner government, and other key stakeholders, including representatives from the beneficiary groups themselves, in all phases of performance measurement. Participation fosters ownership, which is particularly important given the central roles partners play in data collection and use.

Each of these elements or phases is discussed in more detail below.

Phase 1: Formulating objectives

The first step in project performance measurement involves clarifying the project's objectives, by defining precise and measurable statements concerning the results to be achieved, and then identifying the means (i.e., resources and activities/processes) to be employed to meet those objectives. A favourite tool used by the development agencies for conceptualizing a project's objectives and strategies is the project logframe.

The project logframe

The Project Logical Framework, or logframe for short, is an analytical tool (logic model) for graphically conceptualizing the hypothesized cause-and-effect relationships of how project resources and activities will contribute to achievement of objectives or results. The logframe was first developed by USAID in the late 1960s. Since then, it has been adopted by most donor agencies as a project planning and monitoring tool. The analytical structure of the logframe diagrams the causal means-ends relationships of how a project is expected to contribute to objectives. It is then possible to configure indicators for monitoring implementation and results around this structure. The logframe is often presented in a matrix format, for (a) displaying the project design logic (statements the inputs, activities, outputs, purpose and goal), (b) identifying the indicators (and sometimes targets) that will be used to measure progress, (c) identifying data sources or means of verifying progress, and (d) assessing risks or assumptions about external factors beyond project management's control that may affect achievement of results. (See Box 3)

Box 3: Project Design Logical Framework Matrix

Narrative Summary	Objectively Verifiable Indicators	Means of Verification	Important Assumptions
Goal:			
Purpose:			
Outputs:			
Activities:			
Inputs:			

To be used effectively, the logframe should be prepared using a collaborative process that includes different management levels and project stakeholders. Of particular importance is gaining agreement between the donor agency and the partner implementing agency. Although time-consuming, a participatory process is considered essential for building genuine ownership of the project objectives, for testing the logic of the means-ends relationships in debate, and for agreeing on indicators, targets and data collection responsibilities. Most donor agencies encourage broad participation in logframe development, although actual practices may not always live up to policies.

Box 4 provides a generalized version of the analytical structure of the logframe, showing the typical five-level hierarchy used and the types of indicators associated with each level.[8] While most agencies use similar terminology at the lower levels of the logframe hierarchy (inputs, activities, and outputs), there is a confuzing variety of terms used at the two higher levels (called project purpose and goal in this paper). [9] This paper adopts some of the most widely used terms (see Box 4). Note that for some levels, the term (name) used for the hierarchy level itself differs from the term used for its associated indicators, while for other levels the term used are the same.

---

8.   Not all donor agencies use a five-level system; for example, some do not use an activity/process level.

9.   See Annex 1 for a comparison of terms used by different donor agencies.

Box 4: Project Logframe Hierarchy Levels and Types of Indicators

Goal
Impact Indicators

↑

Purpose
Outcome Indicators

↑

Outputs
Output Indicators

↑

Activities
Process Indicators

↑

Inputs
Input Indicators
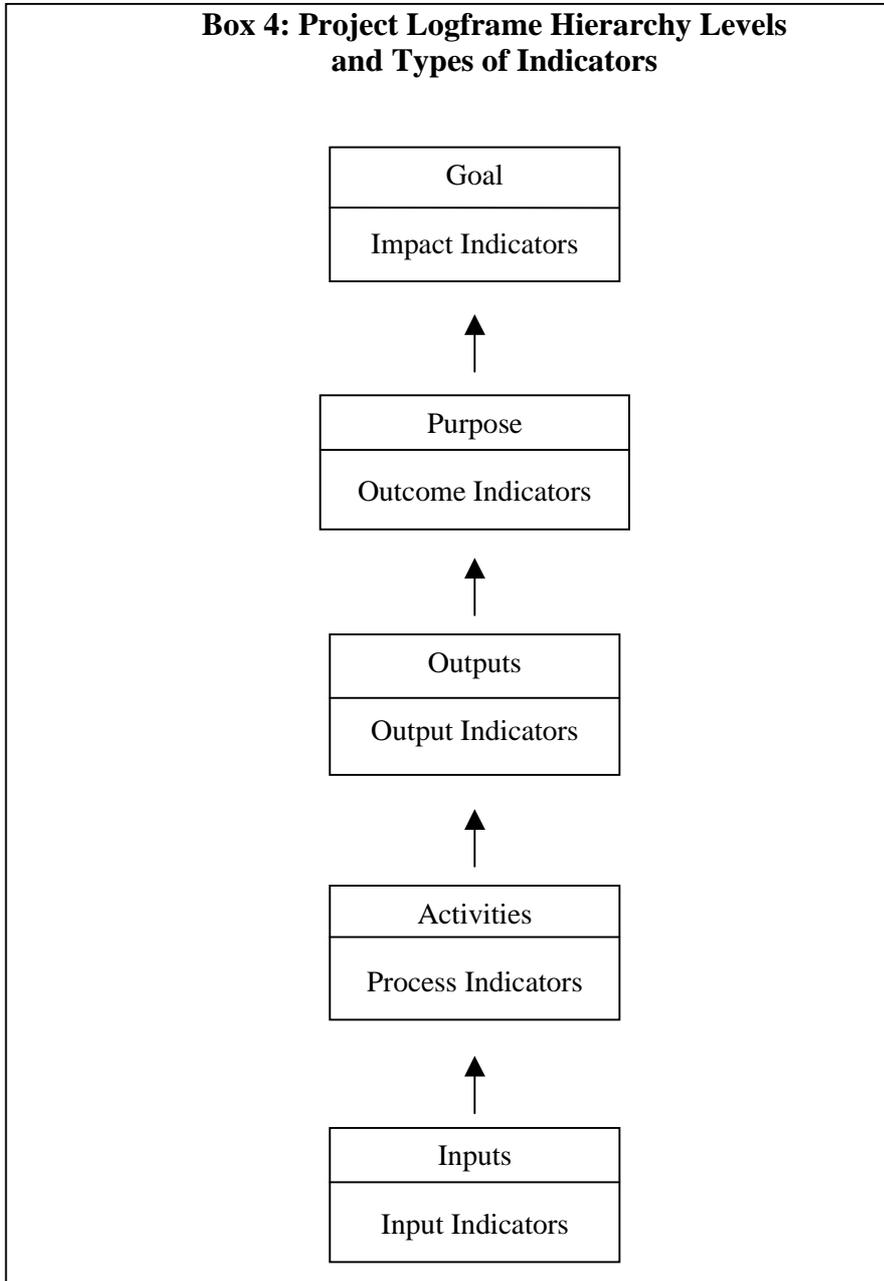
The logframe tool is built on the planning concept of a hierarchy of levels that link project inputs, activities, outputs, purpose and goal. There is an assumed cause-and-effect relationship among these elements, with those at the lower level of the hierarchy contributing to the attainment of those above. Thus, *inputs* are used to undertake project *activities (processes)* that lead to the delivery of *outputs*, that lead to the attainment of the *project purposes (outcomes)* that contributes to a longer-term and broader *project goal (impact)*. The

achievement of each level is also dependent upon fulfilment of certain assumptions in the project's external environment or context that may affect its success.

While there are no standard definitions for the five hierarchy levels that are agreed to or shared by all the development agencies, there are certainly similarities among the definitions used. The definitions below attempt to capture some of these common aspects:

- *Inputs* -- the financial, material and human resources (e.g., funds, staff time, equipment, buildings, etc.) used in conjunction with activities to produce project outputs.

- *Activities (processes)* -- the concrete interventions or tasks that project personnel undertake to transform inputs into outputs.

- *Outputs* -- the products and services produced by the project and provided to intermediary organizations or to direct beneficiaries (customers, clients). Outputs are the most immediate results of activities.

- *Purposes (outcomes)* -- the intermediate effects or consequences of project outputs on intermediary organizations or on project beneficiaries. This may include, for example, their responses to and satisfaction with products or services, as well as the short-to-medium term behavioural or other changes that take place among the client population. Their link to project outputs is usually fairly direct and obvious. The timeframe is such that project purposes or outcomes can be achieved within the project life cycle. Project purposes or outcomes also go by other names -- such as intermediate outcomes or immediate objectives.

- *Goal (impact)* -- the ultimate development objective or impact to which the project contributes -- generally speaking they are long-term, widespread changes in the society, economy, or environment of the partner country. This highest level objective is the broadest and most difficult to attribute to specific project activities. Their timeframe is such that they may not be achieved or measurable within the project life, but only ex post. Other names used at this level include long-term objectives, development objectives, or sector objectives.

The term *results* in this paper applies to the three highest levels of the logframe hierarchy -- outputs, purpose, and goal. Strictly speaking, the lowest levels (i.e., inputs and activities) are not objectives or results, so much as they are means for achieving them.

Difficulty of defining results

Despite attempts to clarify and define three distinct levels of results in the project logframe, reality is often more complex than any logic model. In reality, there may be many levels of objectives/results in the logical cause-and-effect chain. For example, suppose a contraceptive social marketing project provides media messages about family planning and supplies subsidized contraceptives to the public. This may lead to the following multi-level sequence of results:

➢ Contraceptives supplied to pharmacies.
➢ Media messages developed.
➢ Media messages aired on TV.
➢ Customers watch messages.
➢ Customers view information as relevant to their needs.
➢ Customers gain new knowledge, attitudes and skills.
➢ Customers purchase contraceptives.
➢ Customers use new practices.

➢ Contraceptive prevalence rates in the target population increase.
➢ Fertility rates are reduced.
➢ Population growth is slowed.
➢ Social welfare is increased.

What exactly does one define as the outputs....the purpose...the goal? Different development agencies might take somewhat different approaches, varying what they would include in each of the three result categories. Rather than think about categories, it might be more realistic to think, for a moment, about a continuum of results, with outputs at one extreme and goals/impacts at the other extreme. Results along the continuum can be conceptualized as varying along three dimensions -- time, level, and coverage.
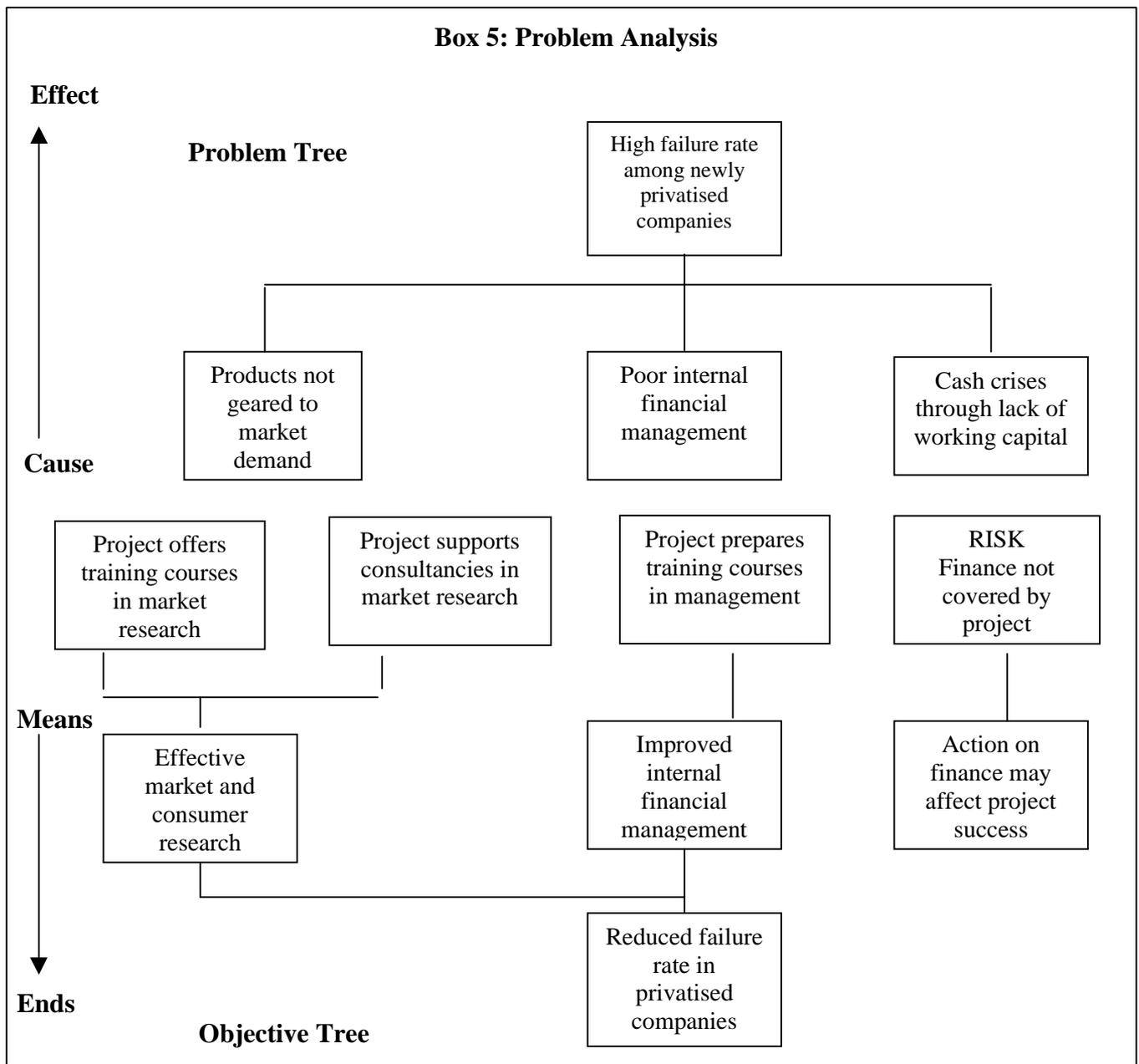
- **Timeframe:** Results range along a continuum from immediate to medium-term to long-term. Outputs are the most immediate of results, while goals (impacts) are the longest-range, with purpose (outcomes) in the middle or intermediate range.

- **Level:** Results also vary along a continuum of cause-effect levels logically related one to the next in a causal chain fashion. Outputs represent the lowest level in the chain, whereas goals (impacts) represent the highest level, while purpose (outcomes) once again fall somewhere in the middle range. Outputs are physical products or services; outcomes are often described in terms of client preferences, responses or behaviors; impacts are generally defined in terms of the ultimate socio-economic development or welfare conditions being sought.

- **Coverage:** A final dimension deals with the breadth of coverage, or who (what target groups) are affected by the change. At one end of the continuum, results may be described narrowly as effects on intermediary organizations or groups, followed by effects on direct beneficiaries or clients. At the other extreme, the results (impacts) usually are defined as more widespread effects on society. Goals tend to be defined more broadly as impacts on a larger target population -- e.g., on a region or even a whole nation, whereas purposes (outcomes) usually refer to narrower effects on project clients only.

However, the nature of goals, purposes, and outputs can vary from agency to agency. Some agencies tend to aim "higher" and "broader", defining their project's ultimate goal in terms of significant improvements in welfare at the national level, whereas other agencies tend to choose a "lower" and "narrower" result over which they have a greater influence. The more resources an agency has to bring to bear to a development problem, the more influence it can exert and the higher and broader it might aim. For example, the World Bank might legitimately define its project's goal (impact) in terms of society- or economy-wide improvements, whereas smaller donor agencies might more appropriately aim at district-level or even community-level measures of change.

Also, if the primary aim of an agency's performance management system is accountability, and managers are held responsible for achieving objectives even at the higher outcome and goal levels, it may be wise for them to select and monitor results that are less ambitious and more directly within their control. If instead, performance management's primary aim is management improvement -- with less focus on strict accountability -- then managers can afford to be more ambitious and define outcomes and goals in terms of more significant results. A challenge of effective performance management is to chose objectives and indicators for monitoring performance that are balanced in terms of their degree of *significance* and *controllability*. Alternatively, agencies need to be more explicit in terms of which levels of results project managers will be held accountable for achieving.

Problem analysis

A useful expansion of the project logframe concept is problem analysis. This is a participatory brainstorming technique in which project planners and stakeholders employ graphic tree diagrams to identify the causes and effects of problems (problem tree) and then structure project objective trees to resolve those problems, represented as a mirror image of the problem tree. Problems that the project cannot address directly then become topics for other projects (possibly by other partners/agencies), or risks to the project's success if no actions are taken. Box 5 provides an illustration of problem and objective trees drawn from the World Bank.

Box 5: Problem Analysis

Effect

Problem Tree

High failure rate among newly privatised companies

Products not geared to market demand

Poor internal financial management

Cash crises through lack of working capital

Cause

Project offers training courses in market research

Project supports consultancies in market research

Project prepares training courses in management

RISK Finance not covered by project

Means

Effective market and consumer research

Improved internal financial management

Action on finance may affect project success

Reduced failure rate in privatised companies

Ends

Objective Tree

Phase 2: Selecting indicators

Once project objectives and the means (strategies) for achieving them have been clarified and agreed upon, the next step is to develop or select indicators for measuring performance at each level of the logframe hierarchy. Performance indicators (simply called indicators hereafter) specify exactly what is to be measured to determine whether progress is being made towards implementing activities and achieving objectives. Whereas an *objective* is a precise statement of what result is to be accomplished (e.g., fertility will be reduced), an *indicator* specifies exactly what is to be measured along a scale or dimension, but does not indicate the direction of change (e.g., total fertility rate). A *target* (discussed later) specifies a particular value for an indicator to be accomplished by a specific date (e.g., total fertility rate is to be reduced to 3.0 by the year 2005).

Types of indicators

The logframe provides the structure around which performance measures or indicators are typically constructed. Different types of indicators correspond to each level of the logframe hierarchy (see Box 4):

Input indicators - measure quantities of physical, human or financial resources provided to the project, often expressed in dollar amounts or amounts of employee time (examples: number of machines procured, number of staff-months of technical assistance provided, levels of financial contributions from the government or co-financiers).

Process indicators - measure what happens during implementation. Often they are expressed as a set of completion or milestone events taken from an activity plan, and may measure the time and/or cost required to complete them (examples: date by which building site is completed, cost of developing textbooks).

Output indicators - track the most immediate results of the project -- that is, the physical quantities of goods produced or services delivered (examples: kilometers of highway completed, number of classrooms built). Outputs may have not only quantity but quality dimensions as well (example: percent of highways completed that meet specific technical standards). They often also include counts of the numbers of clients or beneficiaries that have access to or are served by the project (examples: number of children attending project schools, number of farmers attending project demonstrations).

Outcome indicators - measure relatively direct and short-to-medium term effects of project outputs on intermediary organizations or on the project beneficiaries (clients, customers) -- such as the initial changes in their skills, attitudes, practices or behaviors (examples: project trainees who score well on a test, farmers attending demonstrations who adopt new technology). Often measures of the clients' preferences and satisfaction with product/service quality are also considered as outcomes (example: percent of clients satisfied with quality of health clinic services).

Impact indicators - measure the longer-term and more widespread development changes in the society, economy or environment to which the project contributes. Often these are captured via national sector or sub-sector statistics (examples: reductions in percent of the population living below the poverty line, declines in infant mortality rates, reductions in urban pollution emission rates).

Sometimes a general distinction is made between *implementation indicators* -- that track a project's progress at operational levels (e.g., whether inputs and processes are proceeding according to workplan schedules and within budgets), and *results indicators* -- that measure performance in terms of achieving project objectives (e.g., results at the output, outcome and impact levels). The relative importance of indicator types is likely to change during the life of a project, with initial emphasis placed on input and activity indicators, shifting to output and outcome indicators later in the project cycle, and finally to impact indicators ex post.

While both implementation and results indicators are in this paper considered to be *performance indicators* (just concerned with measuring different aspects of performance), results based management is especially focused on measuring and achieving results.

Also, references are sometimes made to *leading indicators* that are available sooner and more easily than statistics on impact and can act as proxies, or can give early warning about whether impacts are likely to occur or not. Outcome indicators, which represent more intermediate results that must be achieved before the longer-term impact can occur, might be thought of as leading or proxy indicators.

Another type of indicator, often referred to as *risk indicators (also sometimes called situational or context indicators)*, are those that measure social, cultural, economic or political risk factors (called "assumptions" in logframe terminology). Such factors are exogenous or outside the control of the project management, but might affect the project's success or failure. Monitoring these types of data can be important for analyzing why things are or are not working as expected.

Addressing key performance issues

Performance measures may also address any of a number of specific performance issues or criteria, such as those listed below. The exact meanings of these terms may vary from agency to agency. These criteria usually involve making comparisons of some sort (ratios, percentages, etc.), often cutting across the logframe hierarchy levels or sometimes even involving other dimensions. For example:

- Economy -- compares physical inputs with their costs.

- Efficiency -- compares outputs with their costs.

- Productivity -- compares outputs with physical inputs.

- Quality/excellence -- compares quality of outputs to technical standards.

- Customer satisfaction -- compares outputs (goods/services) with customer expectations.

- Effectiveness -- compares actual results with planned results.

- Cost-effectiveness -- compares outcomes/impacts and their costs.

- Attribution -- compares net outcomes/impacts caused by a project to gross outcomes/impacts.

- Sustainability -- compares results during project lifecycle to results continuing afterwards.

- Relevance -- relates project-level objectives to broader country or agency goals.

Different donor agencies tend to place somewhat different emphases on these criteria. Which of the performance criteria are selected generally reflects the primary purposes or uses of the performance management system. For example, if a key aim is to reduce costs (savings), then it is common to focus on cost measures, such as economy and efficiency. If the main objective is accountability, it is usual to focus on output measures, which are directly within the control of project managers. If management improvement is the objective, emphasis is typically on process, customer satisfaction, or effectiveness indicators. Some of these dimensions to performance may present potential conflicts or tradeoffs. For example, achieving higher quality outputs may involve increased costs; efficiency might be improved at the expense of effectiveness, etc. Using a variety of these different indicators may help balance these tensions, and avoid some of the distortions and disincentives that focusing too exclusively on a single performance criteria might create.

Process of selecting indicators

Donor agencies' guidance on selecting indicators generally advises a participatory or collaborative approach involving not only the agency project managers, but also representatives from the implementing agency, partner country government, beneficiaries, and other stakeholders. Not only does it make good sense to draw on their experience and knowledge of data sources, but participation in the indicator selection process can help obtain their consensus and ownership. Given that the responsibility for data collection will often fall to them, gaining their involvement and agreement early on is important.

Steps in the selection process generally begin with a brainstorming session to develop a list of possible indicators for each desired objective or result. The initial list can be inclusive, viewing the result in all its aspects and from all stakeholder perspectives. Next, each possible indicator on the initial list is assessed against a checklist of criteria for judging it's appropriateness and utility. Candidate indicators might be scored against these criteria, to get an overall sense of each indicator's relative merit. The final step is then to select the "best" indicators, forming an optimum set that will meet the need for management-useful information at reasonable cost. The number of indicators selected to track each objective or result should be limited to just a few -- the bare minimum needed to represent the most basic and important dimensions.

Most agencies would agree that the indicator selection process should be participatory, should weigh tradeoffs among various selection criteria, balance quantitative and qualitative indicators, and end up with a limited number that will be practical to monitor.

Checklists for selecting good indicators

There is probably no such thing as an ideal performance indicator, and no perfect method for developing them. Tradeoffs among indicator selection criteria exist. Probably the most important, overarching consideration is that the indicators provide managers with the information they need to do their job.[10] While on the one hand, indicator data should be of sufficient quality to be credible and ensure the right decisions are made, on the other hand they should be practical (timely and affordable).

---

10. How indicator choice relates to uses by different management levels and stakeholder groups is discussed in the next section.

The search for good indicators has prompted the development agencies to devise checklists of characteristics against which proposed indicators can be judged. Although the lists vary from agency to agency in terms of what is emphasized or in the terminology they use to express concepts, there are many overlaps and consistencies among them.

The World Bank suggests that indicators should be relevant, selective (not too many), and practical (for borrower ownership and data collection) and that intermediate and leading indicators for early warning should be included as well as both quantitative and qualitative measures.

USAID's criteria for assessing performance indicators include:

- Direct (valid) -- closely represents the result it is intended to measure.

- Objective -- unambiguous about what is being measured; has a precise operational definition that ensures comparability over time.

- Practical -- data can be collected on a timely basis and at reasonable cost.

- Adequate -- only the minimum number of indicators necessary to ensure that key dimensions of a result are sufficiently captured.

- Reliable -- data are of sufficient quality for confident decision-making.

- Disaggregated where possible -- by characteristics such as sex, age, economic status, and location, so that equitable distribution of results can be assessed.

CIDA's checklist consists of six criteria (posed as questions to consider):

- Validity -- Does it measure the result?

- Reliability -- Is it a consistent measure over time?

- Sensitivity -- When the result changes will it be sensitive to those changes?

- Simplicity -- Will it be easy to collect and analyze the information?

- Utility -- Will the information be useful for decision-making and learning?

- Affordability -- Can the program/project afford to collect the information?

The UNDP's checklist for selecting indicators are:

- Valid -- Does the indicator capture the essence of the desired result?

- Practical -- Are data actually available at reasonable cost and effort?

- Precise meaning -- Do stakeholders agree on exactly what to measure?

- Clear direction -- Are we sure whether an increase is good or bad?

- Owned -- Do stakeholders agree that this indicator makes sense to use?

Box 6 provides additional examples of checklists for selecting performance indicators, from other (non-development) organizations.

Box 6: Examples of Indicator Selection Checklists from other Organizations

Price Waterhouse developed criteria for good performance measures for U.S. government agencies as follows (*Who Will Bell the Cat? A Guide to Performance Measurement in Government*, 1993):

- Objective-linked – directly related to clearly stated objectives for the program.

- Responsibility-linked – matched to specific organizational units that are responsible for, and capable of, taking action to improve performance.

- Organisationally acceptable – valued by all levels in the organization, used as a management tool, and viewed as being "owned" by those accountable for performance.

- Comprehensive – Inclusive of all aspects of program performance; for example, measuring quantity but not quality provides incentives to produce quickly, but not well.

- Credible – Based on accurate and reliable data sources and methods, not open to manipulation or distortion.

- Cost-effective – acceptable in terms of cost to collect and process.

- Compatible – integrated with existing information systems.

- Comparable with other data – useful in making comparisons; for example, performance can be compared from period to period, with peers, to targets, etc.

- Easy to interpret – presented graphically and accompanied by commentary.

In a review of performance measurement (PUMA, *Performance Management in Government: Performance Measurement and Results-oriented Management*, 1994), the OECD concluded that indicators should:

- Be homogeneous.

- Not be influenced by factors other than the performance being evaluated.

- Be collectable at reasonable cost.

- In the case of multi-outputs, reflect as much of the activity as possible.

- Not have dysfunctional consequences if pursued by management.

ITAD (*Monitoring and the Use of Indicators*, EC Report, 1996) developed a popular code for remembering the characteristics of good indictors is SMART:

S – Specific
M – Measurable
A – Attainable
R – Relevant
T – Trackable

There are likely to be tensions or tradeoffs among these various criteria, and a careful weighing of the pros and cons of selecting any particular indicator or set of indicators needs to be made. Some examples:

- A direct indicator of a long-term impact may not be practical or feasible to collect, so a less-than-ideal proxy or leading indicator may have to be accepted.

- Being comprehensive in covering all relevant aspects or dimensions of a result may conflict with the need to limit the number of indicators.

- An indicator selected by a stakeholder in a participatory process may not conform with more conventional or standard indicators that are comparable across projects.

Balancing quantitative and qualitative indicators

Most development agencies agree that both quantitative and qualitative indicators may be useful, and that selecting one or the other should depend on the nature of the assistance program or result. They may be distinguished as follows:

- Quantitative indicators are objectively or independently verifiable numbers or ratios, such as number of people who obtain a hospital treatment; percentage of school children enrolled; output/cost ratios.

- Qualitative indicators are subjective descriptions or categories, such as whether or not a law has been passed or an institution has been established; beneficiaries' assessment of whether a project's services are excellent, satisfactory or poor; or simply a narrative describing change.

Box 7 gives more information on types of quantitative and qualitative indicators, and examples of each.

Box 7: Examples of Quantitative and Qualitative Indicators

Qualitative Indicators: *Illustrative Examples*

    Existence   (a) policy recommendation submitted/not submitted
    (yes/no)   (b) local governance act passed/not passed

    Category   (a) poverty analysed in region "east", "west" or "nationally"
    (e.g., x or y or z)   (b) level of SHD policy focus "high", "medium" or "low"

Quantitative Indicators: *Illustrative Examples*

    Number   (a) number of entrepreneurs trained
                                 (b) number of new jobs created in small enterprise sector

    Percentage   (a) percent of government budget devoted to social sectors
                                 (b) percent of rural population with access to basic health care

    Ratio   (a) ratio of female to male school enrolment
                                 (b) ratio of doctors per 1,000 people

Source: UNDP, *Selecting Key Results Indicators*, May 1999.

Quantitative indicators are readily available in many of the more established "service delivery" sectors of development assistance, such as family planning, education, agriculture, etc. But in other newer or "softer" intervention areas, such as democracy/good governance, policy reform, or institutional capacity-building, the nature of results are such that qualitative indicators and methods may be more appropriate or feasible. The appropriateness of quantitative versus qualitative indicators also depends upon the type of performance issue; for example, quantitative indicators lend themselves to measuring efficiency, while customer satisfaction (subjective opinions) implies using qualitative indicators.

Purely descriptive information, such as a 100 page narrative case study, might not be very appropriate as an *indicator* of change, although it may provide a wealth of useful information about performance. But qualitative information often can be translated into numerical indicators (e.g., by categorizing and counting the frequency of occurrences) that can be useful for monitoring qualitative change. Examples of three common approaches (attitude surveys, rating scales, and scoring systems) are illustrated in Box 8.

**Box 8: Participatory Local Government: Examples of Alternative
Approaches to Measuring Qualitative Change**

Attitude Surveys: Survey respondents may be asked, at different points in time, whether or not they perceive local government to be participatory or not. An improvement in the percentage of people who view local government as participatory -- say from 40% to 65% -- provides a measure that qualitative change is taking place.

Rating Scales: Survey respondents may be asked, at different points in time, to rate their level of involvement on a numerical scale (e.g., from 1 to 10) or according to categories (e.g., very low, low, medium, high, very high). Responses can be presented as averages or as a distribution. For example, between two points in time, the average rating may go up from 2.0 to 7.5 on a 1-10 scale, or the percentage of respondents who consider their involvement to be high or very high may increase from 20% to 50%.

Scoring System: This approach involves devising a scoring system in which values are assigned to observable attributes that are considered to be associated with a particular result. For example, local governments may be considered to be participatory if they have 5 observable characteristics (e.g., holding open public meetings, inviting villages to submit development proposals, etc.). These attributes may be given equal or different weights (depending on their relative importance), and summed into a total score. Then local government units, such as districts, can be scored according to whether or not they have the observable attributes. Over a period of time, improvements may be noted as an increase in average scores of districts, say from 2.4 to 4.0.

Source: UNDP, *Selecting Key Results Indicators, 1999, pp. 6-7.*

Menus of standard indicators

In the search for good indicators, some donor agencies have gone a step further, by providing sector-specific menus of standard indicators. Within broad sector goal or sub-goal areas, projects are categorized into *program approaches* (i.e., similar types of projects sharing common features). For each approach, a set of standard indicators are recommended, or in some cases are required, for project managers' use in measuring and reporting on project outputs, outcomes and impacts.

For example, the World Bank's *Performance Monitoring Indicators*, 1996, offers eighteen volumes of sector-specific technical annexes that provides a structured approach to selecting indicators within each sector/ program area. Most of the sectors follow a typology of indicators based on a hierarchy of objectives and provide a menu of recommended key indicators. USAID has also recently completed similar sector-specific volumes for each of the agency's key goal or program areas that also recommend menus of indicators structured around a hierarchy of objective levels. Danida has taken a similar approach in *First Guidelines for an Output and Outcome Indicator System*, 1998. The guidelines identify standard performance indicators (mostly at the project output level -- to be expanded later to higher outcome levels) that will enable comparable data to be aggregated across similar types of projects to the agency-wide level. Box 9 provides some examples of standard indicators for several types of projects, drawn from Danida's guidelines for the agriculture sector.

```
┌─────────────────────────────────────────────────────────────────────────┐
│                                                                           │
│           Box 9:  Examples of Standard Indicators of Outputs and Outcomes │
│                        From Danida's New Reporting System                 │
│                                                                           │
│   AGRICULTURE                                                             │
│                                                                           │
│      •    Rehabilitation of small scale irrigation schemes:              │
│      •    Number of hectares irrigated per cropping season.              │
│      •    Number of farmers participating.                               │
│      •    Number of participants who are women.                          │
│   Output per hectare of ___ (relevant crop).                             │
│   Farmer Credit:                                                         │
│      •    Number of farmers having formal access to credit.              │
│      •    Number of farmers having formal access to credit through Danish assistance. │
│      •    Number of farmers having or having had a loan.                 │
│      •    Number of these farmers who are women.                         │
│   Source: Danida, First Guidelines for an Output and Outcome Indicator System, September 1998 │
│                                                                           │
└─────────────────────────────────────────────────────────────────────────┘
```

In more decentralized agencies such as USAID, similar menus have been developed, but used only to suggest what are "good" indicators. In Danida the menus are somewhat more mandatory -- to enable aggregation of results to the agency level.

The menu approach holds out the advantage of creating potentially comparable performance data that may facilitate aggregating results across similar projects for agency-wide reporting. However, a problem with the menu approach is that it may appear to emphasize a large number of potential measures, and thus may lead project managers to select too many indicators and over-burden monitoring systems. Moreover, there is always a danger that over time the use of standard indicators may exert pressure to drive project designs into standardized "blueprint" approaches that may not be appropriate in all country contexts. Finally, use of standard indicators provided from the top-down (i.e., from headquarters) may discourage more participatory approaches to selecting indicators.

Indicator selection and different user needs

Choice of performance indicators also explicitly needs to consider the varied information needs of different stakeholder organizations and groups and their various management levels. For example, these would include the field staff and managers in the implementing agency, the project manager, country office director, and senior policy-makers within the donor agency, the project's clients/beneficiaries, officials from the partner country government, etc. The range of measures needs to be sufficiently broad to serve the demands of all these key groups of stakeholders and management levels. Many of these groups will have narrow or partial interests in measures of performance. For example, implementing agency field staff might be most concerned

about indicators tracking whether inputs and activities are proceeding according to plans, whereas unit heads might focus on achievement of output targets. The project's customers/beneficiaries would be most concerned about the achievement of intermediate outcomes and satisfaction measures, which affect them directly. Long-term, socio-economic development impact might be the primary interest of senior policy-makers in the partner country government or in the donor agency, as well as the donor country's parliament and taxpayers.

Within the implementing agency, as the level of management changes, the level of detail of the indicators may change. A field manager, for example, will need to keep detailed records about individual workers, materials purchased, activities completed, etc. on a daily or weekly basis, whereas district or central project managers would require more summary data on a less frequent (monthly or quarterly) basis. The nature of indicators might also shift. At the field level, the priority would be for indicators of resources, costs, and activity milestones, whereas higher management levels would be most interested in efficiency ratios, output and customer satisfaction targets.

The perspectives and indicator needs of different partners and stakeholders may vary as well. Interests of the implementing agency, for example, may be different from those of the donor agency. Implementing agencies tend to be most concerned with indicators of implementation progress, outputs and perhaps with the more project-specific outcomes, but not with broad impact over which they have little control. On the other hand, donor agencies -- especially their senior officials -- are concerned with broad aggregates of social and economic impact. They have use for such information for making strategic policy decisions and also for reporting to their legislative branch and executive oversight agencies concerning the significant development results to which their agencies have contributed.

Senior officials and policy-makers in the partner country governments also have a major stake in impact indicators and data -- much like the donor agencies and their domestic constituencies. But herein lies a potential conflict. If the development impact indicators selected are "driven" by the donor agencies, but each donor agency has different requirements, the amount of duplication and burden on the partner country may be overwhelming. As more and more donors begin to focus on impacts, this problem may multiply unless efforts at harmonization and collaboration among the donors and partner countries increase as well.

It is becoming increasingly clear that performance measurement systems need to be sufficiently comprehensive and balanced in its selection of indicators to cover the needs of all major stakeholders and management levels. For example, focusing only on higher level outcome and impact indicators will not provide an implementing agency with the types of information it needs to implement activities efficiently. Conversely, concentrating only on process and output indicators might result in efficient production of the wrong things, by not providing policy-makers with outcome and impact information they need to make wise policy choices. Similarly, over-emphasis on financial performance may reduce the quality of services or the number of outputs produced. Thus, performance measures should try to cover or balance all major aspects of performance and levels of the objective hierarchy. On the other hand, comprehensiveness may lead to complexity and run counter to the adage to "keep it simple".

An assessment of the flow of information and degree of detail needed by each key stakeholder organization and management level will help clarify the indicators that need to be measured.

Phase 3: Setting targets

Once indicators have been identified for project objectives, the next step often practiced is to devise targets. A target is a specific indicator value to be accomplished by a particular date in the future. Final targets are values

to be achieved by the end of the project, whereas interim targets are expected values at various points-in-time over the life of the project. Baseline values -- which measure conditions at the beginning of a project -- are needed to set realistic targets for achievement within the constraints of resources and time available.

Targets represent commitments signifying what the project intends to achieve in concrete terms, and become the standards against which a project's performance or degree of success will later be judged. Monitoring and analysis of performance then becomes a process of gathering data at periodic intervals and examining actual progress achieved vis-à-vis the target.

Targets may be useful in several respects. They help bring the purposes and objectives of a project into sharp focus. They can help to justify a project by describing in concrete terms what the investment will produce. Targets orient project managers and staff to the tasks to be accomplished and motive them to do their best to ensure the targets are met. They may be the foundation for management contracts clarifying the results for which managers will be held accountable. Finally, they serve as guideposts for judging whether progress is being made on schedule and at levels originally envisioned. In other words, targets tell you how well a project is progressing.

A natural tension exists between the need to set realistic and achievable targets versus setting them high enough to ensure project staff and managers will stretch to achieve them. When motivated, people can often achieve more than they imagined. On the other hand, if targets are unrealistically high and unattainable, confidence and credibility will suffer, and may even set in motion perverse incentives to hide or distort the data.

Any information that helps ground a target setting exercise and ensure its realism is helpful, especially the following:

- *Establishing a baseline*. It is difficult if not impossible to establish a reasonable performance target without a baseline -- the value of the indicator just before project implementation begins. Baselines may be established using existing secondary data sources or may require a primary data collection effort.

- *Identifying trends*. As important as establishing a single baseline value is understanding the underlying historical trend in the indicator value over time. Is there a pattern of change -- a trend upward or downward -- over the last five or ten years that can be drawn from existing records or statistics? Targets should then reflect these trends plus the "value added" that a project is expected to contribute over and above what would have occurred in its absence.

- *Obtaining customer expectations*. While targets should be set on an objective basis of what can be realistically accomplished given certain resources and conditions, it is useful to get opinions from project clients about what they want, need or expect from the project. Customer surveys or consultations can help uncover their expectations of progress.

- *Seeking implementing agency views.* Also important in setting realistic targets is obtaining inputs from implementing agency staff and managers, who will have hands-on understanding of what is feasible to achieve in a given local context. Their participation in the process will also help obtain their agreement to and "ownership" of the targets.

- *Surveying expert opinion.* Another source of information is surveying experts (with technical program area knowledge and understanding of local conditions) about what target is possible or feasible to achieve with respect to a particular indicator and country setting.

- *Reviewing research findings.* Reviewing development literature may help in setting realistic targets, especially in program areas where extensive research findings on development trends are widely available and parameters for what is possible to achieve is already known.

- *Benchmarking.* An increasingly popular way of setting targets is to compare what results similar projects with a reputation for high performance have achieved. These best experiences of other operating units, donor agencies, or partners who have achieved a high level of performance with similar types of projects are called benchmarks. Targets may be set to reflect this "best in the business" experience, provided of course that consideration is given to the comparability of country conditions, resource availability, and other factors likely to influence the performance levels that can be achieved.

Most would agree that setting targets is appropriate for monitoring and judging performance at the lower levels of the logframe hierarchy (e.g., progress in mobilizing resources, in implementing activities, and in producing outputs). Such targets are clearly within the project management's sphere of control. It may also be a useful practice at the intermediate outcome level, which management can reasonably influence although not control completely. However, at the impact level, results are by their very nature affected by many external factors and actors well beyond the project management's control. To the extent that targets tend to imply that project managers are responsible or accountable for achieving them, setting targets at the impact level may be inappropriate or even counterproductive. While impact-level targets may be useful for "selling" a project (competing for funds), the problem is, auditors tend to take them seriously. False expectations may be created. Also, incentives may be produced for managers to distort data or hide negative results rather than report it objectively and transparently.

For impacts, it may be better to simply monitor whether reasonable improvements are occurring in the indicator values rather than to set explicit targets for achievement.

Phase 4: Collecting project performance data

As part of the project planning or design process, indicators are identified, baselines established, and targets set (if appropriate) for each objective. As the project gets underway, empirical observations or data are collected at regular intervals to monitor or measure whether progress is actually occurring.

Generally speaking, project *monitoring* involves the periodic collection of indicator data at all the levels of the project logframe hierarchy. A distinction is often made between *implementation monitoring* -- maintaining records and accounts of project inputs and activities/processes, and *results monitoring* -- measuring results at the output, intermediate outcome and long-term impact levels. A few agencies use the term *performance monitoring* interchangeably with results monitoring, while others use it more broadly covering all levels and types of monitoring. Here, the broader definition is used. The relative importance of monitoring different types of indicator data shifts during the project's life cycle, from an initial focus on implementation

monitoring, to monitoring outputs and intermediate results in the middle years, and finally to the measurement of impact towards the end of the project cycle or ex post.

Implementation monitoring data comes from on-going project financial accounting and field records systems that are maintained routinely by project staff. This information is generally needed frequently (i.e., weekly or monthly) to assess compliance with design budgets, schedules, and workplans. It is used to guide day-to-day operations.

Results monitoring measures whether a project is moving towards its objectives -- that is, what results have been accomplished relative to what was planned (targeted). Information from results monitoring is important not only for influencing medium-term project management decisions aimed at improving the project's performance and achievement of results, but also for reporting to donor agency headquarters. There, it may be combined, aggregated or synthesized with similar data from other projects, and used for making broad policy, program and resource allocation decisions, and also for reporting results to oversight agencies and constituencies.

Monitoring of outputs is the responsibility of project staff and managers and usually involves keeping simple records of numbers of products or services provided and of numbers of clients reached. Output data are collected routinely, usually several times per year. Intermediate outcome monitoring involves obtaining data periodically (e.g., annually) about clients' preferences and responses to the outputs delivered and about their initial effects on clients. While informal consultations with clients might be conducted directly by project staff, often more systematic client surveys, focus groups or other structured rapid appraisal methods are sub-contracted to local organizations, universities or research firms. Monitoring of the project's ultimate impacts -- long-term improvements in the society or economy -- often involve costly population-based sample surveys conducted at the beginning (baseline) and at the end or ex post of the project. Where possible, project-specific efforts might be piggybacked onto household surveys conducted periodically by partner country statistical organizations. This may require financial and capacity-building support by the donor agency to the statistical unit.

As the donor agencies embrace results based management, they tend to shift their focus away from the more traditional implementation monitoring, and give more emphasis to the monitoring of results. Most of the donor agencies reviewed have efforts underway to aggregate project-level results in order to report more broadly on their overall portfolio performance. While these trends towards monitoring higher-order results are desirable, given the historical neglect of measuring at outcome and impact levels, a balance should be sought. The new emphasis on results monitoring should not be at the expense of adequate project monitoring of implementation processes and outputs, over which managers have clearer control and responsibility.

Data collection approaches

Monitoring project performance at the different levels of the logframe hierarchy typically involve different data sources and methods, frequencies of collection, and assignment of responsibility. Good practices involve the preparation of performance monitoring plans at the project's outset that spell out exactly how, when, and who will collect data. Box 10 illustrates a matrix framework tool used by several agencies to record summary information about their monitoring plans.

Box 10: Project Performance Monitoring Plan
(format for recording key aspects of data collection)

Type of Indicator	Indicators and Definitions	Data Sources	Data Collection Methods	Frequency and Schedule of Collection	Responsibility for Data Acquisition
Impact Indicators					
Outcome Indicators					
Output Indicators					
Process Indicators					
Input Indicators					
Risk Indicators					

Some agencies combine plans for data collection with plans for its analysis and use. Key elements typically would include:

- Detailed definitions for each indicator.

- Source and method of data collection.

- Frequency and schedule of data collection.

- Methods of data analysis.

- Identification of those responsible for data collection, analysis and reporting.

- Identification of key users of the performance information.

Donor agency project managers normally have the overall responsibility for ensuring project performance monitoring plans and systems are established. There are several basic options for implementing the data collection (often a mix is chosen, depending on the level of results information and complexity of methods needed):

- *Internal monitoring* - In this case, the project implementing agency is responsible for monitoring. This is the usual option where the implementing agency staff has the capacity and technical skills for performance monitoring (data collection).

- *External monitoring* -- In this case, an external individual or organization (e.g. a consultant, a local organization, a partner government statistical office, etc.) is contracted or supported by the donor agency to independently collect data on results of a project and report to the donor agency. This option is used in cases where data collection is particularly difficult, such as for large or complex projects, or for outcome and impact-level data collection.

- *External support* -- This option combines the above approaches, with the implementing agency responsible for performance monitoring, but with the donor agency providing assistance to help them build their capacity.

As one examines data collection approaches at the various levels of the project logframe hierarchy, certain patterns appear. Box 11 summarizes some of these patterns -- e.g. the typical data collection methods/sources, frequency, and assignment of responsibility -- at each hierarchy level. The location of responsibility for data collection works best if it is placed closely to those who use it. In other words, an organization or management level within an organization may be reluctant to collect data unless it is perceived as directly useful and relevant in its own decision-making processes or tasks at hand. Another pattern, illustrated in Box 12, is the tendency for data collection efforts to become more expensive, time-consuming, and technically complex at higher and higher levels of the logframe hierarchy.
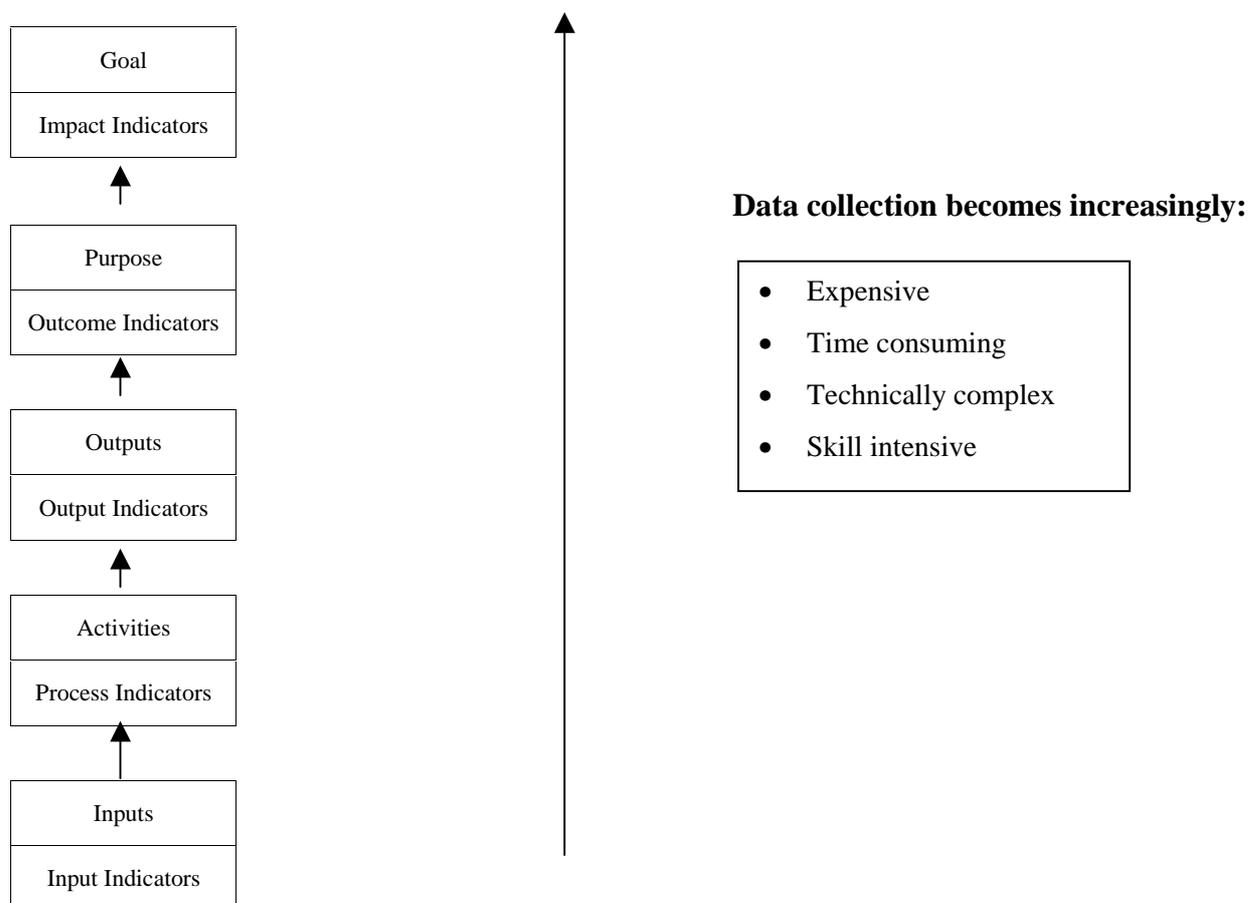
A more detailed discussion follows.

Inputs and process data: Data on inputs and activity processes typically come from project financial accounts and from project management records originating from field sites (e.g., records of resources available and used, of tasks completed, etc.). This level of monitoring is the responsibility of project implementing agency staff and occurs on an on-going basis, with frequent checks to assess compliance with workplans and budget. This type of information is used primarily for day-to-day operations and short-term decisions. The quality of project record keeping in the field can be enhanced by careful attention to design and reporting procedures to ensure validity, replicability and comparability. A good approach is to structure reporting so that aggregates or summaries can be made at intermediate levels -- for example, so that field staff can see how specific villages compare to district averages and improve operations in those villages that are falling behind. While often left out of discussions of project monitoring, a good financial accounting system is needed to keep track of expenditures and provide cost data for analysis of performance issues such as economy, efficiency and cost-effectiveness.

Output data: Data on output indicators (e.g., number of units produced, quality of product or service, number of clients serviced, etc.) also typically originate from project field records maintained by implementing agency staff. Measuring outputs is basically a simple action of counting, but can be complicated in cases where there are many types of outputs whose definitions are not straight forward. Records about clients served (e.g. people attending a clinic, farmers receiving credit) will be more useful in later analysis if their socio-economic characteristics such as age, sex, race, economic status, etc. are kept. Gathering output data are the responsibility of project field staff. The data are aggregated and reported to higher project management levels at regular intervals (e.g. quarterly, bi-annually or annually). Outputs represent the most immediate project results, and their data are useful for short-to-medium term management decisions aimed at improving output quality, equitable distribution to clients, productivity and efficiency, etc.

**Box 11: Characteristics of Project Data Collection Efforts
(By Logframe Hierarchy Levels)**

Type of Indicator	Data Collection Method	Frequency of Data Collection	Organizational Responsibility
Impact Indicators	Censuses and Surveys, National Statistics	Multi-year	Partner Government, Donor Agency
Outcome Indicators	Customer surveys, Rapid Appraisals, Consultations	Annually	Donor Agency, Implementing Agency
Output Indicators	Project Records	Quarterly, Biannually	Implementing Agency
Process Indicators	Project Records	Weekly, Monthly	Implementing Agency
Input Indicators	Project Records, Financial Accounts	Weekly, Monthly	Implementing Agency

**Box 12: Characteristics of Data Collection Efforts
(By Logframe Hierarchy Levels)**

Goal
Impact Indicators

↑

Purpose
Outcome Indicators

↑

Outputs
Output Indicators

↑

Activities
Process Indicators

↑

Inputs
Input Indicators

Data collection becomes increasingly:

- Expensive
- Time consuming
- Technically complex
- Skill intensive

Outcome Data: Measurement of intermediate outcomes typically involves follow-up surveys with project customers/clients on a periodic basis (e.g., annually or whenever there is a need for feedback). These relatively low cost surveys gather information on clients' responses to and satisfaction with project outputs as well as initial effects such as changes in their knowledge, practices and behaviors. Client feedback may be obtained via informal consultations or more systematic approaches such as mini surveys, market research, rapid appraisal or participatory methods. Data should be disaggregated by clients' socio-economic characteristics to facilitate later analysis of equitable distribution of benefits. Outcome data collection may be conducted directly by the project implementing agency if capacity exists or can be developed or it may be sub-contracted to a local organization, university or consultant firm. While relatively uncomplicated and inexpensive, these methods do require some data collection and social science research skills or training beyond regular record keeping and thus should be planned and budgeted for in project design. Outcome data are useful for medium-term management decisions such as those aimed at improving client satisfaction, effectiveness in achieving intermediate results and their equitable distribution.

Impact Data: Measurement of impact generally involves more costly and technically complex population-based sample surveys that can capture more wide-spread and longer-term social and economic improvements, often at the national sector or sub-sector level. Given the long-term nature of these changes (as well as the expense of collecting impact data), it usually only makes sense to undertake such surveys at the project's beginning to establish a baseline and at the end (or even ex post). These efforts are typically beyond the capacity of project implementation agencies to conduct internally.

Where there is a choice, it is usually better to piggyback project-specific impact surveys onto existing national or internationally supported surveys than to create a new data collection facility. If household survey information is already being collected by government agencies or by other donor organizations, it may be less expensive to add on to those efforts than to undertake a separate data collection effort. Project designers need to plan and allow for the costs of collecting impact data; whether they are new surveys or add-ons, there will probably be implications for financial and capacity-building support to statistical organizations.

Simply assuming that existing secondary sources will meet a project's need for impact data without further support may not be justified. Many indicators of impact (e.g. mortality rates, school enrolments, household income, etc.) rely on national surveys or systems of vital statistics. Analysis of project attribution will typically involve comparisons of the situation before and after the project, or in areas covered and not covered by the project. Before data from such sources are chosen as indicators of project impact, the monitoring system designer needs to confirm that the data systems are in place and reliable and valid for the project area and any control groups. Potential problems with using existing data include incomplete coverage of the specific project area, inconsistencies in methods used (e.g. interviewing household members in one survey and household heads in another) or switching techniques (e.g. from measuring actual crop yields in the field to using farmers' estimates). Such problems can invalidate any comparisons intended to show changes in performance. Box 13 gives an example from the World Bank's experience illustrating some of these limitations of survey data. For these reasons, as well as the expense, it may be more appropriate or practical in some cases to rely on using lower level results (e.g. delivery of services, beneficiary/client responses) as proxies or "leading indicators" rather than attempting to measure impact directly.

Impact data are usually not considered to be very relevant by the project implementing agency managers for their own internal decision needs. This is because of its long timeframe (information often not available until after project completion) and its focus on broad socio-economic trends over which the project managers have relatively little influence. Impact data is of most interest to donor agency policy-makers who want this level of performance information for guiding strategic policy and program planning and for resource allocation decisions -- and also for reporting on significant development results achieved to key domestic stakeholders.

```
┌─────────────────────────────────────────────────────────────────────────────┐
│                                                                             │
│           **Box 13: An Example of Limited Comparability of Survey Data**    │
│                                                                             │
│  *Poverty and Household Income*                                             │
│                                                                             │
│  In Uganda, difficulties were encountered in making comparisons between a   │
│  Household Budget Survey carried out in 1989/90 and a later Integrated      │
│  Household Survey in 1992/93. Even under conditions of close supervision    │
│  and rigorous design, small changes in the way in which questions about     │
│  household consumption were put, the layout of the survey form, and         │
│  guidance to enumerators undermined comparability (Appleton, 1996).         │
│  Designers of M&E surveys need to make special provision for comparability  │
│  with existing data from project baseline or national surveys by using      │
│  common survey instruments and methods. The idea that comparisons can be    │
│  made between data collected using different methods is unlikely to pay off. │
│                                                                             │
│  *Source:* World Bank, Designing Project Monitoring and Evaluation, in      │
│  *Lessons and Practices*, OED, June 1996.                                   │
│                                                                             │
└─────────────────────────────────────────────────────────────────────────────┘
```

Contextual Data: For analyzing performance, it is also important to collect data on the project's context – that is, data on exogenous "risk" factors that may affect achievement of intermediate outcomes and especially impacts, but over which the project has no direct control. These factors – be they other partners' interventions, international price changes, armed conflicts or the weather – may significantly affect the achievement or non-achievement of a project's purpose and goal. To the extent they can be foreseen and monitored at reasonable cost, such contextual data can be very useful for explaining project success or failure, and for attributing performance to various causes. See Box 14 for a World Bank example illustrating the importance of collecting contextual data.

```
┌─────────────────────────────────────────────────────────────────────────────┐
│                                                                             │
│           **Box 14: Example of the Importance of Monitoring Risk Factors**  │
│                                                                             │
│  A recent example of a grain storage project in Myanmar demonstrates the    │
│  importance of monitoring risk indicators. During project implementation,   │
│  policy decisions about currency exchange rates and direct access by        │
│  privately owned rice mills to overseas buyers adversely affected the       │
│  profitability of private mills. Management would have been alerted to the  │
│  deteriorating situation had these indicators of the enabling environment   │
│  been carefully monitored. Instead, a narrow focus on input and process     │
│  indicators missed the fundamental change in the assumptions behind the     │
│  project.                                                                   │
│                                                                             │
│  *Source:* World Bank, Designing Project Monitoring and Evaluation, in      │
│  *Lessons and Practices*, OED, June 1996.                                   │
│                                                                             │
└─────────────────────────────────────────────────────────────────────────────┘
```

As donor agencies are increasingly focusing on monitoring impacts (and contextual data), the issue of who is responsible for collecting the data at this level is becoming a growing concern, especially among the NGO community that often implements projects on behalf of the donor agencies. They are feeling under increasing pressure to gather data at this level, while they do not see it as directly related to their implementation-focused concerns. Because of their long-term nature, impacts generally only begin to appear after project completion.

Moreover, the expense and technical difficulty of conducting sample surveys make it well beyond most project implementing agency's capacity and resources. For needed impact and contextual data, it may make more sense for donor agencies to rely on a partner country's national or sector statistical organizations that already conduct surveys regularly and have some capacity -- and to provide additional funding for any project-specific add-ons and capacity-building support as needed.

In addition, it is becoming increasingly important that donor agencies working within a partner country on similar development objectives co-ordinate their requirements for impact data. As more and more development agencies become increasingly results-focused and begin establishing performance measurement systems that encompass broader outcomes and impacts, the potential for over-burdening the capacities of partner country institutions with duplicating data collection efforts increases. Harmonization of data collection requirements for impact (and contextual data) is thus fast becoming a priority needing attention.[11] Co-ordinated support to partner country statistical organizations and cost-sharing arrangements may also help reduce the cost burden of expensive surveys on individual donor agencies.[12]

Criteria for selecting data collection methods and sources

Selection of a data collection method and source can be important in terms of the data's quality -- e.g., its validity and reliability, but also in terms of its practicality or feasibility given cost and time constraints. For example, if information is needed on farmers' use of fertilizer, this might come from either extension agents' records or alternatively from a production survey. Choosing the survey may result in greater statistical validity and reliability of data, but on the other hand using the extension agents' records might mean the data can be collected more frequently and at lower cost. Thus, there are obvious tradeoffs to be considered among criteria for selecting a data collection method and source. The selection process should aim at balancing the need for data to be of a sufficient quality to be credible among its intended users (e.g., to make appropriate decisions or to convince auditors and other stakeholders), versus its implications for cost and timeliness.

Key criteria used for selecting data collection methods include the following measurement and practical issues.

Measurement Issues:

Validity - A measurement is valid to the extent that it represents what it is intended and presumed to represent, and has no systematic bias. Do the data mean what we think they mean? Does the measurement technique indeed measure what it purports to measure? Validity may be affected by conceptual or by technical errors of all kinds. For example, an indicator that is selected just because it happens to be available, may not be considered meaningful or valid by others. But even if the indicator itself is valid, technical errors during the data collection process may make measurement invalid (e.g., unrepresentative sampling, non-sampling errors, etc.)

---

11. This may be viewed as part of a broader need for donors to harmonize their development activities within a partner country, e.g., by employing new joint sector assistance modes.

12. Some donor agencies have on occasion contracted with evaluation or research firms from outside the partner country to conduct sample surveys and analyze impact of projects. While these firms may have the needed skills to conduct credible data collection and analysis efforts, such an approach does not build partner country capacity and moreover is less likely to result in donor coordination and cost-sharing.

Reliability - refers to the stability and consistency of the data collection process (over time and from place to place), so that measured progress reflects real changes rather than variations in data collection procedures, methods or techniques. A measurement is reliable to the extent that, repeatedly applied to a given situation, it consistently produces the same results if the situation has not changed between applications. For example, an IQ test would be unreliable if it is administered twice to the same person, whose intelligence hasn't changed, but produces two different scores. As is the case with data validity, measurement error can compromise the reliability of data. Unreliability may originate from several sources, such as interviewer error or improper implementation of sampling procedures.

Practical issues:

Timeliness - the data can be collected frequently (routinely) enough and is current enough to inform management's decision-making processes. Some methods can be more quickly implemented and thus are more suitable if data are needed urgently or at frequent intervals.
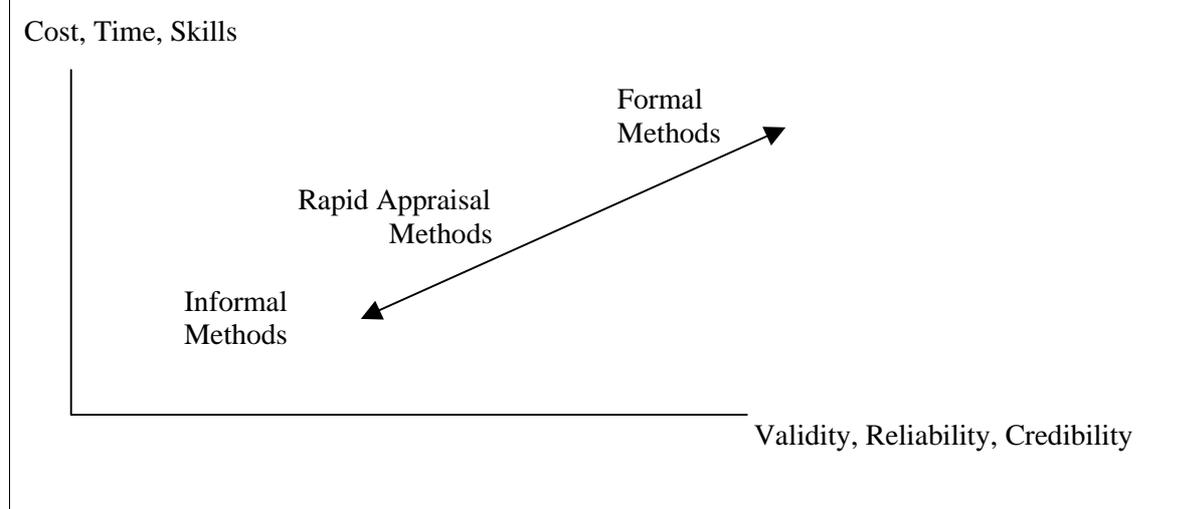
Costs - the cost of a data collection method is a practical issue that must be considered in choosing among methods. Some methods -- e.g., large sample surveys -- are expensive and may not be affordable within budget constraints.

Formal versus informal methods

Data collection methods may be viewed as arrayed on a continuum, with very informal methods, such as casual conversations or unstructured site visits, on one side, and highly formal methods, such as censuses and sample surveys, on the other side. While informal methods may be quick and cheap, they may not be as credible with decision-makers as the more formal, rigorous methods. Whereas formal methods have comparatively high reliability and validity (and therefore credibility), they typically are expensive and time-consuming and require extensive technical skills. Moreover, in developing country settings where technical capacities are frequently limited, they have often been plagued with implementation problems.

In-between the two extremes lie rapid appraisal methods. Some of the most popular rapid appraisal methods include key informant interviews, focus groups, community interviews, direct observation, and mini-surveys. Being neither very informal nor fully formal, they share the properties of both, and that is both their strength and their weakness. Box 15 illustrates general tradeoffs between these types of methods. Which method is most appropriate for a given situation depends on the felt need for high quality and credible data versus practical constraints of cost, time and skills available.

<div style="border:1px solid">

Box 15: Tradeoffs Among Different Types of Methods

Cost, Time, Skills

                            Formal
                            Methods

Rapid Appraisal
       Methods

Informal
Methods

                                  Validity, Reliability, Credibility

</div>

Generally speaking, obtaining impact level data will typically require more formal methods, such as population-based sample surveys, whereas rapid appraisal methods are often useful for gathering outcome data.

Specific advantages of rapid appraisal methods include their relatively low cost, quick turn-around time, and flexibility. They can provide in-depth information about an issue, process or phenomenon. Moreover, they can be learned relatively quickly, making them more suitable for participatory approaches. Their shortcomings include limited reliability and validity, frequent lack of quantitative information from which generalizations can be made, and possibly less credibility with decision-makers than the formal methods. Some rapid appraisal methods (e.g. direct observation, mini surveys, community interviews) tend to produce more quantitative information than others (e.g. focus groups, key informant interviews); however the quantitative information is often not representative, as are the more formal methods.

Quantitative versus Qualitative Methods

A related issue is choice of quantitative versus qualitative methods. While there appears to be some favoritism or emphasis on using quantitative methods for performance indicators and monitoring, most donors note the utility of both types of information and the need to balance both. Some donors stress that quantitative methods should be utilized wherever possible for performance monitoring, but recognize that qualitative methods will be necessary where techniques are not yet sufficiently developed to provide quantitative measures.

Rather than think of quantitative and qualitative methods as two contrasting or opposing options, it may be more helpful to think of a continuum with varying degrees of quantification. On the quantitative extreme are measures that involve continual, equal-interval scales with true zero points (such as GNP per capita, infant mortality rates, school enrolment rates, etc). At the qualitative extreme are data that can be captured only by descriptive narrative. In the middle are data for which the frequency of various events can be counted and categorized, and perhaps even rank-ordered. A good deal of innovative work is now going on in the middle

area. For example, much of the performance data being collected on policy reform, institutional strengthening, and customer feedback are measured on some type of ranked (ordinal) scale. Such scales, when clearly operationalized, provide an example of how more subjective information can be usefully and effectively quantified. See Box 16 for an illustration drawn from USAID for monitoring progress in legal policy reform projects.

The quantitative-qualitative debate is not necessarily an either-or question with some predefined answer. Choice of more quantitative or qualitative indicators (and their associated methods) involve tradeoffs -- often between practicality (cost/timeliness) on the one hand, and objectivity, validity (directness) and reliability (comparability) on the other. Whether quantitative or qualitative, performance measures need to permit regular, systematic and relatively objective judgements to be made about change or improvements, and requires adequate data comparability over time.

Box 16: Measuring Stages of the Legal Reform Process

This example illustrates how qualitative information can be translated into quantitative through use of scales. Drawn from USAID guidance for measuring progress in democratic legal reform, this type of methodology could have applications for any type of policy reform. The reform process is broken down into key stages (each of which is fully defined in the guidance):

Stage 1 - Interested groups propose that legislation is needed on issue
Stage 2 - Issue is introduced in the relevant legislative committee or executive ministry
Stage 3 - Legislation is drafted by relevant committee or executive ministry
Stage 4 - Legislation is debated by the legislature
Stage 5 - Legislation is passed by full approval process needed in legislature
Stage 6 - Legislation is approved by executive branch (where necessary)
Stage 7 - Implementing actions are taken
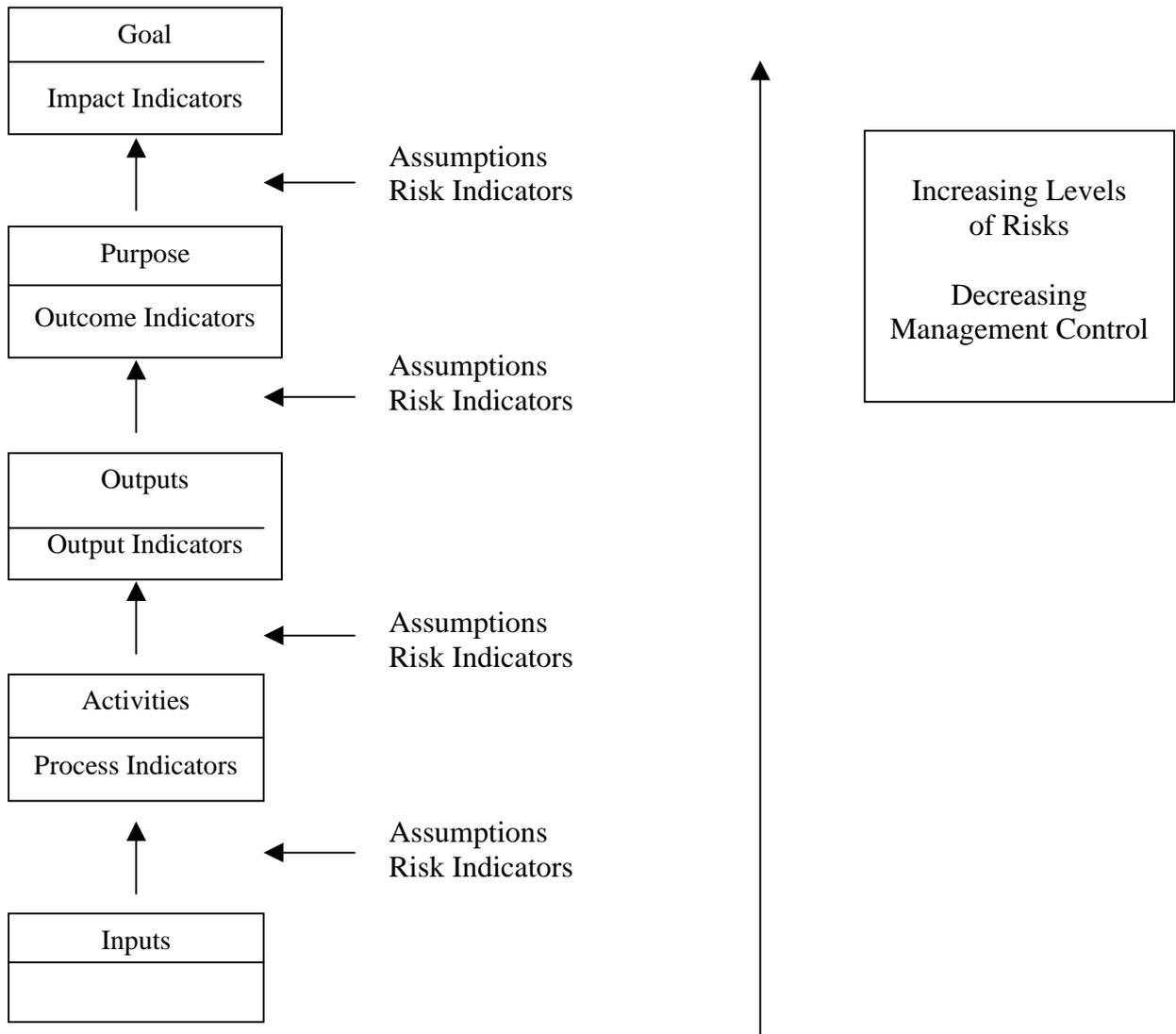Stage 8 - No immediate need identified for amendments to the law

Each stage could be given a value of 1 (or weighting could be applied if there is reason to assign differential importance to each stage). Scoring method: Performance is reported by presenting the highest stage (milestone) passed during that year. Performance targets are set by planning what stage is expected to be achieved in the coming years for each reform being promoted.

Source: USAID*, Handbook of Democracy and Governance Program Indicators,* April 1998*.*

Risks and management control at different hierarchy levels -- implication for accountability

As one moves up the different levels of the project logframe hierarchy, there are more and more external influences and increasing levels of risk, and consequently decreasing degrees of management control. (See Box 17.)

Box 17: Relationship Between Project Management Control and Risk Factors (by logframe hierarchy levels)

Goal
Impact Indicators

↑   ←   Assumptions
Risk Indicators

Purpose
Outcome Indicators

↑   ←   Assumptions
Risk Indicators

Outputs
Output Indicators

↑   ←   Assumptions
Risk Indicators

Activities
Process Indicators

↑   ←   Assumptions
Risk Indicators

Inputs

Increasing Levels
of Risks

Decreasing
Management Control

Inputs, activities and outputs are mostly within the direct control of project managers. Management accountability at these levels is usually not controversial. With intermediate outcomes, one begins to move away from internal management control into the external environment. For example, client responses and preferences are not directly controllable by project managers. Nevertheless, often managers still have strong influence over results at this level and may be willing to be held accountable. However, at the impact level (i.e., broader, long-term sector development improvements) external risk factors are generally quite influential, and numerous other interventions, development partners and actors may also be affecting results. Thus, it becomes increasingly more difficult and unjustified to hold project managers accountable for achieving specific results/targets at this level. Moreover, because of the many external factors and actors influencing these broad impacts, demonstrating attribution can prove difficult.

It may be useful to think of a continuum of results with different amounts of controllability and intrinsic value (significance). At the one extreme are the outputs (goods and services), which are highly controllable but of little significance. At the other extreme, the results may be of great importance, but be well beyond the control of project managers to achieve.

Balancing monitoring needs at different results levels

At what level should the focus of performance monitoring be placed? Concentrating on just one level of result may have unintended, even dysfunctional, consequences. For example, concentrating only on the output level may result in "doing the wrong thing well". Concentrating only on high-level impacts may lead to lack of basic monitoring information about project activities and services, and result in poor implementation.

The answer appears to lie in taking as comprehensive and balanced an approach as is possible, within limits of what's reasonable and practical. Developing a more comprehensive performance monitoring system that recognises the need for performance information at various levels is least likely to lead to distortions. Moreover, as already discussed, different stakeholder groups and management levels will have varying interests in these levels of results, so satisfying everyone means having a comprehensive system.

Being comprehensive in monitoring at all results levels, however, may take more effort and resources than can normally be handled by project implementing agencies. Clearly, monitoring of project inputs, activities, outputs, and perhaps even intermediate outcomes closely related to the project should be an internal project management function and responsibility (and moreover is relatively simple and low cost). However, at impact levels it may make more sense to co-ordinate joint performance monitoring efforts with other donors and partners also working towards the same development objective within a country.

A note of caution, however, is that while trying to be comprehensive, one should equally avoid building project performance monitoring systems that are too complex -- and that will detract from other staff functions such as implementing activities or conducting complementary evaluations.

Phase 5: Reviewing and reporting project performance

Most of the donor agencies reviewed are now requiring project management to conduct annual reviews assessing and reporting project performance monitoring data.[13] These reports are usually mandatory for all

---

13.   An exception is USAID which no longer requires any reporting to headquarters at the project level, but does require reporting at the broader country program level (see next chapter).

larger projects. Analysis of project performance monitoring data may address a broad variety of performance issues, both implementation and results-oriented. However, emphasis is now mostly on results (compared to previous project monitoring reports that focused on implementation activities). Moreover, the current focus is chiefly on effectiveness in achieving results as planned (targeted). Some agencies also require a comparison of budgeted versus actual expenditures for each set of activities used to generate an output.

These periodic assessments of performance monitoring data by project management help alert them to performance problems, which may lead directly to taking corrective actions (where causes are fairly straight-forward) or may signal the need for more in-depth studies/evaluations focused on understanding specific performance shortfalls and recommending actions. Analysis of performance monitoring data generally gives a good indication of *whether* performance is on-track, compared to some standard or target, but may not adequately explain *why* or *how* performance fell short of (or exceeded) expectations. This is a special domain or role for evaluation.

In addition to serving as a management tool at the project level to aid in project decision-making, these annual project performance reports provide agency headquarters with building blocks for aggregating and analyzing performance of the overall project portfolio. Such agency-wide analyses can be useful in policy formulation, program strategy development, resource allocation processes, and in annual agency performance reporting to external stakeholders.

A number of donors who have traditionally had most project performance information assessed and reported at project completion (via project completion reports) are now moving to require increasingly formalized and rigorous analyses of performance on an annual basis for on-going projects. These assessments or reviews are prepared at regular intervals (usually annually) during implementation to provide "real time" early warning information and feedback about performance geared to making project improvements. These reports are also serving to provide more consistent, comparable and regularly available data on project performance and results to headquarters where it can be aggregated or summarized across the agency's portfolio in annual agency performance reports. For example, DFID is asking project managers to annually assess progress towards achievement of the project purpose, and to assign performance ratings in reports called *output-to-purpose reviews*. CIDA requires project managers to prepare similar *annual project progress reports* comparing its actual results to date compared to its expected results, and assigning performance ratings. AusAID also has a new requirement that project managers periodically prepare *project monitoring briefs* that assess, rate, and report performance for on-going projects. These periodic assessments are usually self-assessments by project management. Because of their timing during implementation, sometimes assessments or performance ratings are based on *likely* rather than *actual* achievement of the longer-term outcome and impact targets. Most of these agencies also continue to require project completion reports that, because of their timing, can better report on actual outcomes achieved (accomplishment of project purpose), although they may still not be able to report on actual long-term impacts or sustainability, which may only show up sometime after project completion.

Summaries of donor approaches to periodic assessments of project performance

Brief summaries are provided below of different donor agencies' requirements for regular project-level reports reviewing performance monitoring data.

- CIDA's annual progress project report (APPR) system tracks active projects valued over $100,000. These reports, prepared by project managers, follow a consistent format. They compare a project's actual expenditures versus budget, actual results versus targets, identify lessons learned, and assign

performance ratings. The key focus of these reports is on comparing results actually achieved to-date with expected results. Shortfalls from what was expected are signals of possible problems that need to be discussed, reviewed, and resolved. These reports are also used to prepare the agency's performance report to Parliament. In addition, CIDA project managers are required to prepare a self-assessment at project completion, called project closing reports, which is the primary mechanism for results reporting at corporate levels. These reports not only assess whether a project has achieved its results but also make an effort to explain why or why not, and convey more completely the full nature of a project, including its relevance, sustainability and cost-effectiveness.

- DFID has a similar system in which managers of larger projects (over L500,000) are being asked to annually assess progress towards achievement of the project purpose, and to assign performance ratings. These substantive reviews are called output to purpose reports. They focus attention towards project purpose and away from the usual concentration on activities and outputs. The format of these reviews is tabular, with its origins in the logframe matrix, and is geared to management decisions. Emphasis is on improving performance as opposed to simply auditing projects. Risks and risk management are assessed and corrective action is taken where necessary. These annual self-assessments of on-going projects are becoming increasingly rigorous and formalized. Still, data quality control remains an issue, and consideration is now being given to the need to build random quality checks into the system. DFID also requires project completion reports for larger projects that assess what the projects have achieved and lessons learned. Project performance information (beginning with project ratings) is now being combined with financial information in a central automated database, and will be used for agency-wide portfolio performance reporting.

- AusAID has just installed a new system of activity monitoring briefs (AMBs) for the periodic assessment of performance and assignment of ratings for on-going projects at 3-12 month intervals, depending on the riskiness of the project. The briefs are considered a basic monitoring tool to assist in the management of project activities, and also will serve to provide standardized reporting on project information. Preparation of the briefs is the joint responsibility of country desks and posts, and is based on consultations with partners and stakeholders. Each project is rated against four criteria – appropriateness of design, achievement of objectives, quality of management, and sustainability of outcomes. Project performance ratings from the AMBs are stored in a centralized automated database that – along with project financial data – will be used for annual agency performance reporting across the portfolio and for accrual budgeting.

- Danida has begun a new system requiring country offices (embassies) to report annually on project outputs and outcomes using sets of comparable indicators and standard reporting formats for each of Danida's key program approaches. This system is primarily geared to headquarters' needs for external reporting to domestic stakeholders on Danida's overall results achieved, and not as an internal management tool. A first annual report based on this new system was recently published.

- The World Bank requires performance reporting on a regular basis (typically semi-annually) for all projects, submitted by the project task managers after field supervision missions. The reporting format, known as Form 590, requires self-ratings of progress in implementation and in achievement of expected results. They incorporate key indicators of development results and provide space for explanation and justification of the ratings. The information from these forms are entered into a central database and analyzed in the Bank's annual report on portfolio performance (ARPP). In addition, the Bank requires assessment and reporting of performance at project completion. Self-evaluations called implementation completion reports (ICRs) are prepared by Bank managers and borrowers (partners) within six months

of the final disbursement of the Bank loan. These reports review issues encountered during implementation, achievements, and lessons learned. They evaluate results (benefits) and where applicable re-calculate the economic rate of return. They also include an outlook section based on the borrower's plan for the operational phase of the project, including performance indicators to monitor the project and measure its ultimate development impact. ICRs rate the project's outcome (taking into account the project's relevance, efficacy, and efficiency), sustainability, institutional development impact, and also the quality of performance by both the Bank and the borrower. The World Bank's Operations Evaluation Department (OED) reviews and validates (or adjusts where necessary) the performance ratings of all IRCs. For about 25% of completed projects, OED also conducts independent, field-based performance audit reports to verify the project's implementation and results. They also include an initial assessment of the project's longer-term effects. OED maintains a database on project performance based on ICRs and performance audits, which it uses as building blocks for assessments of sector, country, or special issues/themes. OED also has for many years conducted and published annually an aggregated analysis of Bank-wide project performance based on ICRs and audits completed during the year, called annual review of evaluation results.

- USAID's and the UNDP's approach to assessing and reporting results is based on the concept of broader country programs, not individual projects (discussed in the next chapter).

These annual project performance reports and completion reports are generally considered an integral part of the performance measurement system. Their relationship to evaluation -- dealt with much more fully in a later chapter -- is not always that clear cut. These reports do review, analyze, and make judgements about performance issues and results, and are thus evaluative in nature. However, a few key distinctions are mentioned here.

- The performance monitoring reports are self-assessments by project managers, whereas evaluations are typically conducted by larger evaluations teams, often comprised of external evaluators that can provide an independent judgement about project performance. However, trends towards more participatory forms of evaluation in some agencies may make this less of a distinction.

- The performance monitoring reports are mandatory for larger projects and thus provide a reasonably complete coverage of the overall project portfolio, whereas evaluations are often conducted on a much more selective (i.e., occasional or optional) basis for projects of particular interest or concern.

- Performance reports involve relatively straightforward presentations of actual results achieved vis-à-vis expected results, actual expenditures data vis-à-vis budgets, management self-ratings for various aspects of project performance, etc. Typically performance data and/or ratings are presented in standard, comparable formats that can be easily entered into databases and summarized across the portfolio. They are meant to provide consistent types of information covering a broad range of performance issues and results, but without great depth of analysis. Evaluations, on the other hand, usually are less standardized in format and content, following individual scopes of work. Moreover, they tend to focus on fewer performance issues but analyze them in greater depth.

- Performance monitoring reports focus mostly on whether or not results were achieved as planned, whereas evaluations can better explain why and how they were achieved or not. In other words, evaluations seek to analyze and understand the project's context and factors influencing performance, both internal (within management control) and external (beyond management control).

- Because of timing as well as the need to use more rigorous analytical methods, assessments of long-term impact, attribution of impacts to specific donor activities, cost-effectiveness, and sustainability are

49

performance issues that evaluations can often address better than annual performance monitoring reports.

Types of performance issues often addressed in performance monitoring reports

Analysis of project performance monitoring data in annual project progress and project completion reports may be directed at assessing a number of performance issues, some of which address implementation performance and others of which address results achieved. Some of these issues draw on monitoring data that cut across the logframe hierarchy levels, or that relate different aspects to each other (e.g., costs to outputs). Some of the most commonly raised performance issues are listed in Box 18.

Box 18: Key Performance Issues

- *Economy* - the relationship between costs and physical inputs (i.e., an organization is economical if it is purchasing inputs as cheaply as possible).
- *Efficiency* - the relationship between costs and outputs (example: cost per kilometer of road built).
- *Productivity* - relationship between inputs and outputs (example: number of demonstrations handled per extension worker).
- *Excellence* – producing high quality outputs (example: percent of units produced that meets technical standards).
- *Equity* – the extent to which needy or disadvantaged sub-populations have equitable access to results (example: percentage of students attending project schools who are female).
- *Customer satisfaction* - how well project outputs correspond to client preferences (example: per cent of clients satisfied with clinic services delivered).
- *Effectiveness* - the extent to which results -- outputs, outcomes, or impacts – are being achieved as planned (targeted).
- *Attribution* - the extent to which observed outcomes and impacts can be attributed to a particular project. That is, separating the net outcomes/impacts caused by a project from gross outcomes/impacts.
- *Cost-effectiveness* - the relationship between project costs and net outcomes/impacts attributable to the project.
- *Sustainability* - the capacity for results to extend beyond the formal life of the project.
- *Relevance* - the continued appropriateness of a project's results - to the needs of the target population, to the partner country's national development goals, and to the development agency's corporate-level objectives.

While performance monitoring is fairly well positioned to directly assess the relatively straight-forward nature of the performance issues of economy, efficiency, productivity, excellence, equity, customer satisfaction, and effectiveness, alone it may not be sufficient to adequately analyze the more complex performance issues of attribution, cost-effectiveness, sustainability, or relevance. First of all, the timing is such that longer-term impacts and sustainability typically cannot actually be assessed in monitoring reports, other than in terms of likelihood. Moreover, analysis of these issues can be technically demanding and require expertise, time and effort beyond the typical project manager's or implementing agency's capacity. While performance monitoring data should be drawn upon in the assessment of these issues, often supplemental, more in-depth efforts and additional fieldwork are required to address them adequately -- such as special evaluation studies designed to address them (see chapter on role of evaluation).

This variety of performance issues reflects that there are a number of dimensions to performance. Some may even involve potential conflicts or tradeoffs. For example, achieving higher quality outputs may involve increased costs; efficiency might be improved at the expense of effectiveness, etc. It is usually a good idea to address as many key performance issues as possible (given resource constraints and the need to keep performance measurement and analysis systems reasonably simple), because exclusive focus on one aspect or the other may create unintended distortions and disincentives.

A few of these performance issues are addressed in detail below.

Analyzing effectiveness in achieving results

One of the most common types of analysis of performance monitoring data examines *effectiveness*. This usually involves straightforward comparison of actual results achieved against targets (i.e., planned or expected levels of results to be achieved by specific dates). Most agencies are getting on the bandwagon and setting explicit targets. However, using targets tends to imply management control and accountability for achieving them. While targets may be very appropriate for outputs, and probably also for intermediate outcomes, their appropriateness for the goal/impact level might well be questioned, given management's very limited sphere of control or influence at this level. Alternative approaches to assessing effectiveness might be to simply look for continuous improvements or positive movement towards objectives and then make retrospective judgements whether this amount of progress was reasonable, given the constraints and external influences that occurred. Benchmark comparisons of a project's progress in achieving results with that of "high-performing" projects of a similar type is another popular analytical technique for assessing effectiveness. Comparisons with the effectiveness of similar types of projects implemented by other country offices or donor agencies can be made. However, the differences among country conditions, resource availability, capacities and starting points make such cross-country comparisons of progress difficult. For example, one couldn't reasonably expect the same rate of improvement in countries that are at very different stages of development.

Analyzing results achieved in relation to costs

While not all agencies' performance measurement systems as yet emphasize *efficiency and cost-effectiveness* issues, others already place priority on such concerns. Some development agencies, such as the World Bank, discuss results based management in terms of achieving results as efficiently and cost-effectively as possible -- that is, in terms of ensuring resources flow to areas that make the most impact[14]. CIDA requires reporting of

---

14.   World Bank, *Performance Management in the World Bank*, presentation for the DAC Workshop on Performance Management and Evaluation, October 1998.

cumulative budgets versus actual disbursements to date by activity sets (groups of activities associated with producing a particular output) as part of the annual project progress reports.

In some other agencies the emphasis on measuring results has thus far not taken much explicit account of the costs involved in achieving those results. In a number of cases, legislation on performance based budgeting will soon require agencies to begin to analyze results in relation to their costs. There are numerous constraints facing agencies in doing this, which will be discussed in subsequent chapters.

Analysis of project efficiency (outputs in relation to their costs) and cost-effectiveness (higher-order results in relation to their costs) both require that good financial/expenditure records are kept that can be related to particular project outputs, outcomes, or impacts. Efficiency can then be analyzed fairly straightforwardly in terms of cost per unit of output. Performance may be assessed by comparing actual with planned unit costs, or in terms of finding low cost options. Efficiency criteria may involve tradeoffs with quality of output concerns, which any analysis should weigh.

Cost-effectiveness analysis has the additional problem of having to deal with the attribution problem, since we really want to know about costs in relation to the net outcomes or impacts directly resulting from the project activities, and not from external factors. Comparisons of cost-effectiveness across similar projects can tell managers which approaches can achieve a given outcome or impact at least cost.

Project rating systems

A common approach to analyzing and judging project performance in performance monitoring reports is the use of project rating or scoring systems. Rating systems are defined as instruments or structures for judging performance and results of projects by using a standardized set of criteria for such judgement and a standardized rating scale. Project rating systems are able to combine in one measure, or rating, different performance issues or criteria. Moreover, they have the advantage of being able to compare and aggregate ratings across different types of projects. There has been a growing interest in and use of rating systems in the donor agencies, due to the increased attention to performance and accountability, and the growing need to be able to obtain an aggregate overview of a sometimes large and complex portfolio of projects and programs.

Criteria and rating scales used vary considerably from agency to agency. Rating scales have either an even or an odd number of categories. For example, an even-numbered scale might be: "highly satisfactory", "satisfactory", "unsatisfactory", and "highly unsatisfactory". Criteria may be sector-specific or crosscutting (such as effectiveness, efficiency, or sustainability). Each criteria are usually scored separately, then averaged to give an overall project rating. The sets of criteria used may be relatively simple or complex. Box 19 provides an example of a project rating system -- from AusAID.

Performance data for rating projects usually come from performance monitoring systems or sometimes from evaluations. The ratings or judgements themselves are often self-assessments by project management made at specific times during implementation or at project completion, but they may also be made (or validated) by independent evaluations or performance audits.

Lessons learned about practices for establishing an effective rating system are provided in Box 20.

Rating systems may have many uses at different management levels. At the project level, they are intended as a tool to alert managers to emerging performance problems that need corrective action or further study. But, perhaps one of the most important attributes of rating systems is that they can be aggregated across dissimilar

types of projects, and thus provide a mechanism for comparing and reporting performance and results at various higher organizational levels (e.g., country programs, sector programs, department and agency-wide levels).

Box 19: Example of a Project Rating System – AusAID

AusAID has recently (June/July 1999) issued new guidance for *Activity Monitoring Briefs* (AMBs). The AMB is a basic monitoring tool to assist in the management of project activities. Also, in the context of results-based reporting and accrual budgeting, AMBs will provide a basis for standardized reporting on activity performance against four well-defined quality criteria:

- Appropriateness of objectives and design
- Achievement of objectives
- Professionalism of management
- Sustainability of outcomes

A project is rated against each of these criteria on a five-point rating system:

     5 = best practice
     4 = fully satisfactory
     3 = satisfactory overall
     2 = marginally satisfactory
     1 = weak

The AMB rating is an indicative assessment of activity (project) performance and likely sustainability at the time of assessment. Preparing AMBs are the joint responsibility of country Desks and Posts, and should involve consultations with partners, stakeholders and beneficiaries. AMBs are to be prepared every 3 to 12 months, depending on the assessed level of risk of the project.

The performance ratings for individual projects are entered and stored in AusAID's central, automated Activity Management System database. This facilitates aggregation, assessment and reporting of performance at various organizational levels, including at the country and overall portfolio and sector levels (i.e., for each of AusAID's key goal areas, called Key Result Areas).

Source: AusAID, Circulars on *Performance Information Under Accrual Budgeting* and on *Activity Monitoring Brief: Operational Guidelines* (June/July 1999)

Box 20: Effective Practices for Establishing Project Rating Systems

- Rating scales should be clearly differentiated to facilitate choices. Controversial, extreme ratings such as "failure" should be avoided. It's best to have an even number of categories to prevent the tendency of raters to choose the middle category.

- Using numeric codes along with corresponding descriptions of categories will facilitate later analysis.

- Instituting checks and balances, such as scrutiny of ratings by management and audits, will improve the system's reliability.

- Ratings should be consistent with the purpose and timing of the rating. Ratings used for supervision should emphasize criteria related to implementation issues, processes, outputs and short-term outcomes. Ratings used for project completion reporting and evaluation should focus on the longer-term and include criteria related to the degree of achievement of objectives and targets, and issues such as effectiveness, impact, and sustainability.

- Adequate and visible use of ratings in decision-making and in assessing and reporting aid effectiveness provides motivation to staff to apply the rating system.

- Rating systems require an organizational culture that values performance and results, learning from successes and failures, and is willing to be open, transparent and accountable.

- A balance needs to be sought between simplicity on the one hand versus comprehensive coverage of relevant aspects on the other. Too much emphasis on comprehensiveness could lead to complex systems that are difficult to operate.

- Rating systems should be an integral part of the management system and integrated into the project cycle. A sound basis for rating needs to be established at the planning stage with clear definitions of objectives, targets, assumptions and measurable indicators. Ratings should be based on adequate information from monitoring, reviews, project completion reports and evaluations.

- Rating should be the responsibility of the officer best positioned. Ratings of progress and short-term results are best prepared by project managers, whereas rating of performance and impact should be carried out by evaluators at project completion.

Source: Rating Systems in Aid Management - Executive Summary, prepared for the DAC Working Party on Aid Evaluation, DCD/DAC/EV(96)9, September 1996.

Analyzing attribution

Attributing outcomes and impacts to specific project interventions generally falls under the domain of evaluation rather than performance monitoring reports. However, it is raised here because it is a growing issue of concern among various oversight agencies and auditors who review the project performance measurement and reporting systems of donor agencies.

The question of *attribution* arises in relation to project outcomes and, even more so, impacts. How can the effect of a project on an outcome or impact be distinguished from all other possible influences? In other

words, the concern is with separating the *net* outcome or impact directly attributable to (caused by) a specific intervention, from the *gross* observed outcome or impact that may also have been caused by other factors (e.g., other partners' activities, weather, political instability, policy environment, etc.) Thought of another way, the issue is to try to compare what outcomes/impacts would have occurred in the absence of the project – to what actually occurred with the project. Attribution is related to accountability concerns in the minds of many auditors. If a donor or its implementing agency mangers are considered *responsible* for achieving a particular outcome/impact target, it must be demonstrated as *attributable* to them as well, at least in part.

Usually the complexity of assessing attribution, especially at the impact level, takes it beyond the normal realm of performance monitoring reports (with its typically straightforward analysis), and makes it the subject of special evaluations. Such evaluations, however, can usefully draw on results data from monitoring systems, if they exist (see Box 21).

Box 21: Approaches to Analyzing Attribution
Using Performance Monitoring Data

Performance monitoring data may tell us whether we have reason to believe whether our activities are contributing to important objectives, but it cannot provide scientific proof of a cause-effect relationship. However, one can make a good case of "plausible association", much like a lawyer would build a case in a court of law.

Confidence in the cause-effect relationship can be increased by doing one or more of the following:

1. Pick objectives (e.g., intermediate outcomes) that are not far removed from the project management's level of responsibility and control -- where the causal relationships are fairly direct and obvious
2. Focus on the logic of the strategy, particularly on identifying and monitoring intermediate outcomes as critical linkages between assistance activities and the impact level (i.e. by building a "chain of evidence")
3. Monitor the "critical assumptions" or risk factors that may also influence the outcome/impact. Where possible, eliminate alternative explanations for what may have caused the result.
4. Use disaggregated data, if available, and compare results (outcomes, impacts) among assisted and non-assisted populations.
5. Examine historical trends in the outcome/ impact data -- before and after the project -- to see if rates of change accelerated
6. Supplement the quantitative performance monitoring data with qualitative evidence about linkages -- for example, by drawing on expert opinion, key informants, focus groups, case studies, etc.

Usually, performance monitoring data documents the observed gross changes in outcomes or impacts, which are then analyzed in relation to planned targets. Analysis of attribution typically goes beyond the normal realm of performance monitoring reports. More in-depth assessments or evaluations are usually needed.

Shortcomings of the project logframe approach

Because of its focus on single projects, the logframe approach to project management and measurement tends to lead to well-designed and carefully monitored stand-alone projects. A donor agency's portfolio of projects within a country may therefore be unrelated to one another, rather than be aligned (i.e., combined or concentrated) into more coherent program strategies aimed at achieving a few significant development objectives.

Another problem with the project logframe approach is also inherent in its basic project orientation. It tends to lock managers into a particular "blueprint" set of project activities that may or may not be effective in reaching results. It ignores the fact that many development projects are fundamentally experiments or pilot efforts with considerable risk and uncertainty regarding the ability to achieve results. Although mid-course project corrections or re-design is possible in theory, often constraints exist that make it difficult to "let the managers manage" to achieve results. An approach that allows managers greater flexibility and authority over choice of activities and outputs might be more in line with the new results based management philosophy of devolving authority over resources and activities as well as accountability.

Performance monitoring systems focused on individual projects may be limited in their time-horizons. If there are no follow-on or post-project monitoring arrangements made, it will be difficult to track longer-term impacts or sustainablility of results.

Also, because the project logframe is by its nature focused on individual projects by a single agency, it can not adequately or comprehensively address how the joint activities of numerous development partners working together might all contribute to achievement of a shared development objective.

See Box 22 for more critiques of the project logframe approach.

The next chapter reviews donor experiences with performance measurement systems developed for broader *country programs* – defined as sets of related projects or other activities (by multiple development partners) that contribute to the same development objective within a country, usually at a national sector level. By focusing on broader programs, some of the shortcomings of the project logframe approach can be overcome.

Box 22: Limitations of the Project Logframe Approach

While it is a powerful tool, the project logframe has a number of shortcomings in the way it is often conceptualized or applied. For example:

- The logframe often assumes that necessary resources will be provided, and thus does not adequately reflect the financial environment in its indicators.

- The real sequence of cause-effect relationships may not fit neatly into the logframe's simple 5-level hierarchy. For example, there may be several intervening intermediate outcomes between outputs and final impact.

- The levels of objectives do not necessarily identify which level of results managers will be held responsible or accountable for achieving. The implication is usually that they are responsible for all levels. This is usually not realistic at the goal (impact) level and may create false expectations.

- All too often, agency staff prepare the logframe as a routine formality before project design submission, rather than taking the time to involve stakeholders and partners in the process to gain agreement on objectives and indicators.

- The analysis of risks is often weak, with little formal assessment of the external contextual factors and of other actors (e.g., partner country governments, civil society organizations, other donor agencies) that will also influence achievement of an objective. Ideally, indicators measuring key assumptions or risks should be identified and monitored.

- By detailing project activities/ processes, the logframe tends to support a "blueprint" design approach that may not allow managers enough flexibility to shift easily from less promising to more effective activities.

- The focus of the logframe on a single project by a single agency does not apply well to newly emerging joint sector program assistance modes -- in which the partner country and numerous donor agencies join forces and harmonize their activities to achieve a shared sector objective.

IV. PERFORMANCE MEASUREMENT IN THE DEVELOPMENT CO-OPERATION AGENCIES

-- *The Country Program Level* --

This chapter examines how some donors have responded to the perceived shortcomings of the project level approach to performance measurement by shifting to a more strategic, country level focus. It explores how results-oriented measurement processes, techniques and tools, such as strategic planning (including results frameworks) and performance monitoring systems, are being applied at the country program level. Under this newer approach, the country operating units (i.e., country field missions or offices) of the donor agencies are encouraged to work in harmony with other development partners to achieve shared development objectives within a particular country setting. Ideally, donor operating units apply country-level strategic planning and performance monitoring systems and tools that have been developed jointly or in partnership with other donors, under the leadership of the partner country government. These systems are to be used for guiding and harmonizing the donors' country assistance programs based on their comparative advantages and special interests.

The country program approach is a much more comprehensive and strategic approach to performance management and measurement than the project approach. It focuses on a significant development objective within a country, usually at a broad sector or sub-sector level (but may also focus on a crosscutting objective such as gender equality), and measures the collective performance of numerous projects and non-project activities all contributing to that objective. Thus, the unit of analysis is not a single project but a broader country program that typically includes many projects and other activities implemented by different donor agencies and other partner organizations over a relatively long time period.[15] Performance measurement frameworks and systems developed at the country program level are thus comprehensive, long-term, multi-activity and multi-agency endeavors.

Unlike the project approach, which puts equal weight on monitoring all levels of the logframe hierarchy, and may even traditionally have favored implementation monitoring, the country program approach puts the higher-level development objectives at center-stage. The shift from individual projects to programs also implies a different timeframe dimension, freed from the confines of a single project's life cycle. By focusing on a country level development objective and the intermediate outcomes needed to achieve that objective, the timeframe now becomes longer-term, outliving the comings and goings of individual project activities.

USAID pioneered this approach during the mid-1990s, abandoning its previous focus on projects and moving towards more strategic and results-oriented country programming approaches, as part of its broader "reengineering" reforms. The UNDP's new results based management system also adopts this model.

---

15. This distinction between a project and a program was somewhat unique to USAID (who first introduced the approach), until recently. Now a number of donor agencies have adopted a similar definition of program.

The World Bank has also recently initiated more strategic approaches to planning, implementing, and measuring performance of development programs at the country level as part of its "comprehensive development framework" approach. (See Box 23.)

Box 23: The Comprehensive Development Framework

The Comprehensive Development Framework (CDF) proposal was introduced by World Bank President James D.Wolfensohn in January 1999, in response to alarming statistical trends showing the war on poverty is being lost in much of the developing world. The CDF calls for a new way of doing business, not only for the World Bank but also for all partners engaged in the development enterprise – e.g., partner developing country governments, their civil society institutions (media, NGOs, etc), private sectors, and other donor agencies. It calls for fundamental transformations in the way development operations or processes are now being conducted in partner countries in order to improve development effectiveness – that is, to achieve better development results. Four overarching themes comprise the CDF:

1. **Long-term, holistic strategy.** Developing national visions/strategies for development that are long-term, comprehensive and integrated -- balancing attention between macroeconomic/financial and structural/social development objectives.

2. **Ownership & Participation.** Putting the partner country in the driver's seat in formulating and implementing the development strategy, while involving broad participation from society in the process.

3. **Partnership.** Changing donor practices to support country-led aid co-ordination, to link their individual assistance strategies with the country's overall development strategy, and to practice selectivity (division of labor) based on comparative advantages.

4. **Focus on Results.** Fostering a results-orientation in partner country and donor organizations, involving up-front strategic planning (clarifying expected outcomes, indicators and targets), collection of performance information (monitoring and evaluation), and the use of performance information in management accountability, learning, and decision-making processes

These concepts are not totally new ideas. In fact, there is a considerable consensus emerging within the international development community that they should be encouraged in development co-operation practices. However, putting the four principles together into an integrated, synergistic package or approach is new.

To test the CDF approach and learn lessons about effective practices, the World Bank undertook pilots in 12 developing countries beginning in March 1999. Several Bank reports, based on information from pilot self-monitoring systems, have already reviewed the CDF country pilots' initial progress in applying the principles, including discussions of constraints and tensions that challenge their implementation, and of some promising practices.

Sources: CDF Progress Report to the World Bank's Executive Board, May 2000; OED*, 1999 Annual Review of Development Effectiveness; Comprehensive Development Framework – Report on Country Experience*, September 2000

One advantage of using this country level approach, from the perspective of a donor agency's operating unit, is that it clarifies the few significant development objectives that the unit seeks to contribute to in the partner developing country. It helps identify what program strategies are necessary and sufficient to achieve a significant development objective, and then enables the unit, working collaboratively with other development partners, to sort out their individual comparative advantages, and responsibilities/ contributions to the overall strategy. This can help the operating unit to better align (focus and concentrate) their assistance activities into those few program strategies for which they have taken responsibility, rather than just have a diverse portfolio of seemingly unrelated projects. The country development objectives and intervention strategies selected by a unit usually have to be in line with the donor agency's overall corporate goals/priorities and areas of comparative advantage. Moreover, they should directly contribute to the goals articulated in the partner country's national development vision or strategy. This new approach requires some fundamental changes in the way development partners conduct their in-country development activities compared to past practices. In particular, it calls for adherence to principles of country ownership/leadership in development processes, participation by civil society and the private sector, partnership or close collaboration among donors, and a shared results-orientation by all partners.

Individual project activities tend to be less well defined in this approach, allowing for more flexible designs and implementation, rather than rigid "blueprint" approaches. Moreover, in some donor agencies (e.g., USAID) projects no longer require headquarters approval. Instead, authority is delegated to operating units in the field so they can shift course mid-stream if results monitoring information indicate certain activities are not working well. The World Bank has also been experimenting with sector-wide approaches supported by flexible programmatic investment instruments, such as sector investment and maintenance loans and adaptable program loans (APLs). Moving from projects to a full-scale sector-wide approach involving the pooling of multi-donor and government finances has also been piloted in some countries.[16]

The country program level approach puts a premium on partnerships and more collaborative approaches, since achieving a strategic, long-term development objective is clearly dependent on the activities of many development partners -- e.g., various donor agencies, the NGO community and private sector, and of course the partner country government. Donors should facilitate putting the partner government "in the driver's seat" in leading and co-ordinating these processes. In some cases this will have to involve building the partner country's capacities and internal demands for performance measurement and management. Some of the tools developed for country program level strategic planning and performance monitoring should be particularly well suited to new modes of development co-operation based on joint multi-partner, sector-wide programs in which investments and activities are harmonized to achieve shared country development objectives.

The process of performance measurement at the country program level involves the same five elements or phases as is the case at the project level:

- Formulating objectives.
- Selecting indicators.
- Setting targets.
- Monitoring performance (collecting data).
- Reviewing and reporting performance data.

---

16.   World Bank, *Expenditure Framework and Partnership*, OED Working Paper Series, summer 2000.

Many of the concepts, issues, processes and tools used for country-level performance measurement are similar to those used at the project level, and thus do not have to be repeated in detail here. This chapter attempts to highlight some of the differences and advantages as well as disadvantages that are specific to the country-level perspective. Practicing performance measurement at both project and country program levels, and effectively linking them, is also discussed.

Generally speaking, there is less experience with performance measurement at the country program level than at the project level. A number of constraints have limited the extent to which donors have been willing or able to actually adopt these harmonized practices, despite "policies" that in principle support greater participation and partnership. The approaches presented below are thus to some extent "ideal" or best practices. Actual practice has probably been less participatory and more "donor-centered" than it should be, especially given its strategic nature.

Phase 1: Formulating objectives

Results frameworks

A relatively new and useful tool that is being used by some donor agencies for strategic planning and performance measurement at the country program level is the *results framework* (also called program logframes, performance frameworks, etc.). A results framework is a graphic display of the strategies necessary and sufficient for achieving a significant or strategic *development objective* in a partner developing country, usually at a sector or sub-sector level. Thus, it embodies the development hypotheses underlying strategies for the achievement of the objective. The results framework relies on objective tree concepts, and diagrams the logical cause-effect relationships between activity outputs at the bottom, intermediate results or outcomes in the middle, and the strategic development objective at the top.

Thus, the results framework is similar to the project logframe in that they are both based on logic models or objective tree concepts linking different levels of objectives via cause-and-effect relationships. Results frameworks differ by explicitly including the outcomes of many related projects (and non-project activities), not just one project. Moreover, results frameworks concentrate on the higher-order results levels (not activity inputs or processes) and are more flexible in allowing several levels of intermediate outcomes (in between outputs and development objectives).

The individual elements (levels) usually included in a results framework include:

- *Development objective* - is a significant or strategic development objective for a country that a group of development partners may share and be jointly responsible for achieving. It is generally a long-term and widespread development change, usually at the national sector level. It is the highest-order result in the framework. In other words, a results framework has only one strategic objective -- at its pinnacle. Its nature is such that partners are jointly accountable for its achievement. It is somewhat similar in concept to the impact level in the project logframe.

- *Intermediate outcomes* - (also called intermediate results, program outcomes, or simply outcomes) are the intermediate level results that contribute to achieving the strategic development objective. They are the next lower levels on the logical cause-effect chain. Taken as a group, the intermediate outcomes are those results both necessary and sufficient for achieving the development objective. They are medium-term results usually affecting a specific target group of intermediate organizations or beneficiaries (i.e., not society-wide). They are the direct consequence of outputs from sets of project or non-project

activities. Intermediate outcomes in the results framework are close in concept to that of outcomes in the project logframe, except in the results framework, all intermediate results necessary and sufficient to achieve the development objective must be included, not just those related to one project. Thus, they represent the combined outcomes of a number of related projects and non-project activities, rather than just one project. Usually, partners will divide responsibilities for achieving intermediate outcomes among themselves, based on selective interests, capacities or comparative advantages. Thus, a single donor agency's operating unit might take responsibility (and accountability) for achieving one or more of these intermediate outcomes, although it is also conceivable that several partners might take joint responsibility for one.[17]

- *Outputs* - underneath the intermediate outcomes in the cause-effect hierarchy are the sets of outputs from related project and non-project activities. Each set of activity outputs are those that together are necessary to achieve a particular intermediate outcome. Activities are not necessarily limited to those of donors and the partner government but may also include actors from civil society and the private sector.

The results framework defines the cause-and-effect linkages between the outputs, the intermediate results, and the development objective and also identifies the critical assumptions that must hold true for the results at various levels to be achieved. In other words, the results framework embodies the development hypotheses implicit in the partners' harmonized strategies for achieving the shared development objective.

Box 24 illustrates the general structure and levels of a results framework diagram. Note that there may be more than one level of intermediate outcome in the cause-effect chain (providing more flexibility than the rigid hierarchy levels found in project logframes). At the bottom of the diagram are the sets of activity outputs (possibly from more than one project) designed to meet each intermediate result.[18] Box 25 provides a hypothetical example of a results framework (limited to development objective and intermediate outcome levels). Note the useful practice of making it explicit in the diagram which agencies or partners are responsible for which intermediate outcomes.
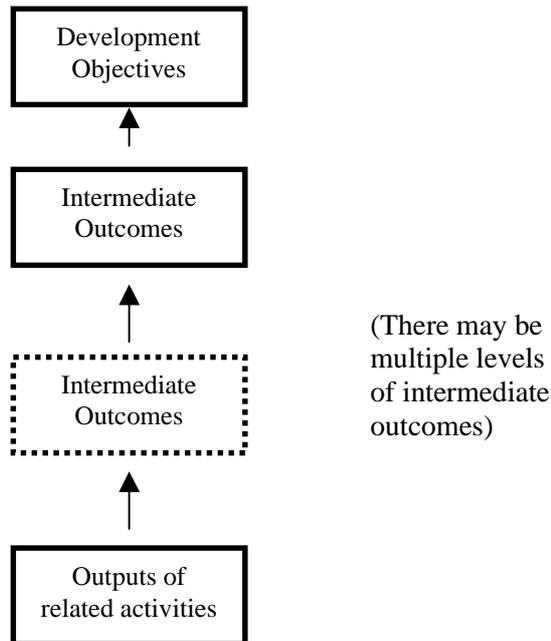
Results frameworks are useful as strategic planning and management tools. They can help a donor agency's operating units at the country level to clarify the significant development objectives they seek to contribute towards, in collaboration with partners, and to develop a logical plan or program strategy for their part in achieving them. Over time, operating units can begin to align (focus and concentrate) their activities within coherent programs that support the specific intermediate outcomes for which they have taken responsibility.

---

17. It is at this level of result (intermediate outcomes) that it probably makes most sense to hold individual partners, such as donor operating units, accountable. While they are not as completely within their control as outputs, and are of course subject to some external risk factors, they are close enough in the cause-effect chain for project managers to exert considerable influence. On the other hand, achievement of the development objective might best be viewed as a shared or joint accountability among partners.

18. Project inputs and processes are generally not detailed in results frameworks, which are meant to focus on results, not implementation processes. Moreover, some agencies (e.g., USAID) do not even routinely include the output level in their results frameworks.

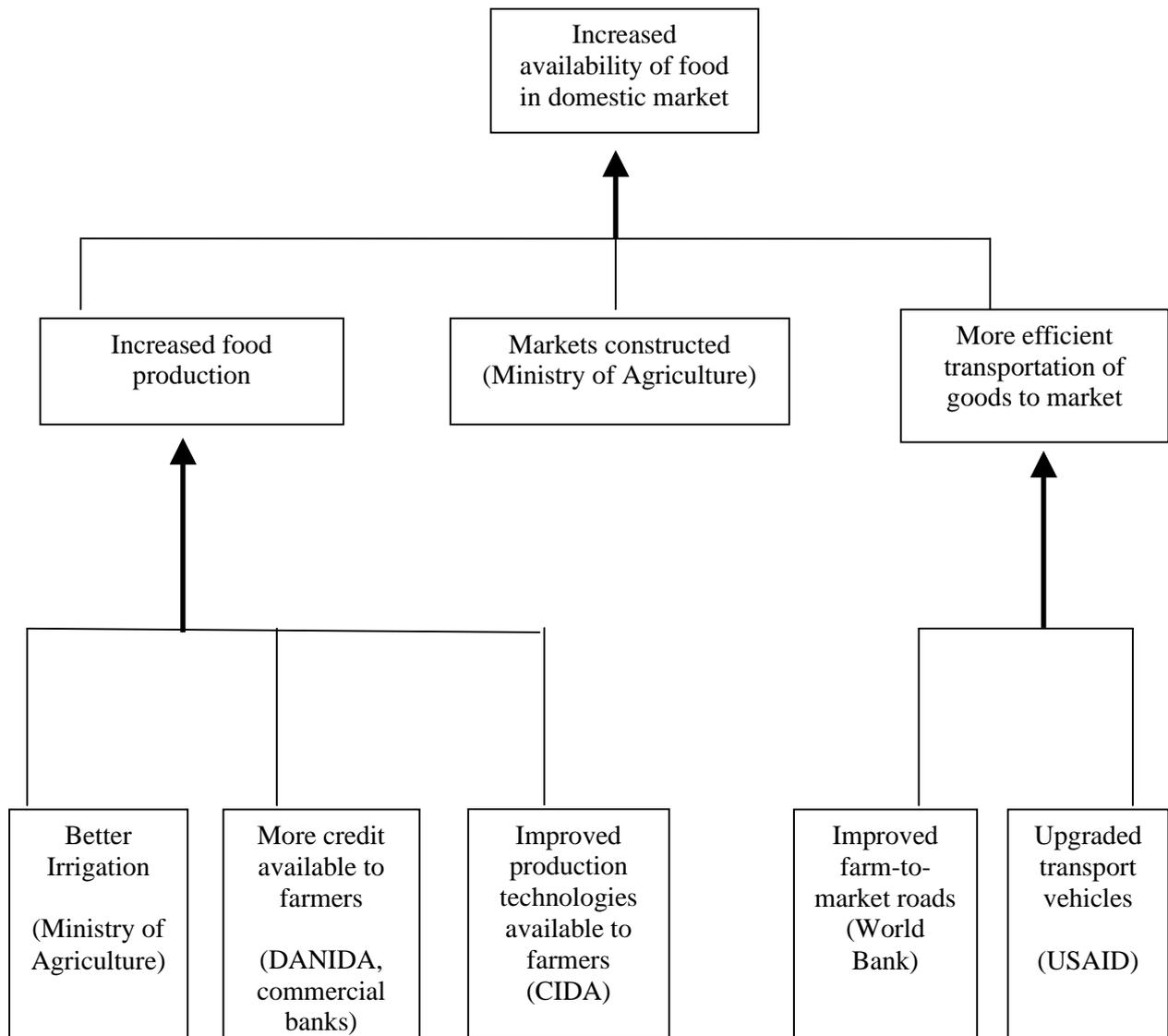**Box: 24: General Structure of a Country Program Results Framework**

(Hierarchy Levels)

```
        ┌─────────────────┐
        │  Development    │
        │  Objectives     │
        └─────────────────┘
                 ▲
                 │
        ┌─────────────────┐
        │  Intermediate   │
        │  Outcomes       │
        └─────────────────┘
                 ▲
                 │
        ┌ ─ ─ ─ ─ ─ ─ ─ ─ ┐
          Intermediate           (There may be
        │ Outcomes        │      multiple levels
                                 of intermediate
        └ ─ ─ ─ ─ ─ ─ ─ ─ ┘      outcomes)
                 ▲
                 │
        ┌─────────────────┐
        │  Outputs of     │
        │  related        │
        │  activities     │
        └─────────────────┘
```

The results framework is also a performance measurement tool -- providing a structure for measuring and monitoring progress towards the achievement of those results for which the unit is responsible. Performance data from the monitoring system is then used to alert managers when actual results are not as planned, and require adjustments to be made in relevant projects and activities. In addition, the design of a results framework provides an opportunity for an operating unit to work with its development partners to build consensus and joint ownership around shared objectives. Results frameworks also function as effective communication tools because they succinctly capture the key elements of a strategy for achieving an objective (i.e., program intent and content).

Box 25: Hypothetical Illustration of a Country Program Results Framework

(Development Objective and Intermediate Outcomes)

```
                          ┌─────────────────────┐
                          │     Increased       │
                          │ availability of food│
                          │  in domestic market │
                          └─────────────────────┘
                                    ▲
        ┌───────────────────────────┼───────────────────────────┐
        │                           │                           │
┌────────────────┐        ┌────────────────────┐      ┌──────────────────┐
│ Increased food │        │ Markets constructed│      │  More efficient  │
│  production    │        │(Ministry of        │      │ transportation of│
│                │        │  Agriculture)      │      │ goods to market  │
└────────────────┘        └────────────────────┘      └──────────────────┘
        ▲                                                       ▲
 ┌──────┼──────────────────┐                           ┌────────┼────────┐
 │                         │                           │                 │
┌──────────┐ ┌──────────┐ ┌──────────┐       ┌──────────┐ ┌──────────┐
│ Better   │ │More credit│ │Improved  │       │Improved  │ │Upgraded  │
│Irrigation│ │available to│ │production │       │farm-to-  │ │transport │
│          │ │farmers   │ │technologies│      │market roads│ │vehicles  │
│(Ministry of││          │ │available to│      │(World    │ │          │
│Agriculture)││(DANIDA,  │ │farmers    │       │Bank)     │ │(USAID)   │
│          │ │commercial│ │(CIDA)     │       │          │ │          │
│          │ │banks)    │ │          │        │          │ │          │
└──────────┘ └──────────┘ └──────────┘       └──────────┘ └──────────┘
```
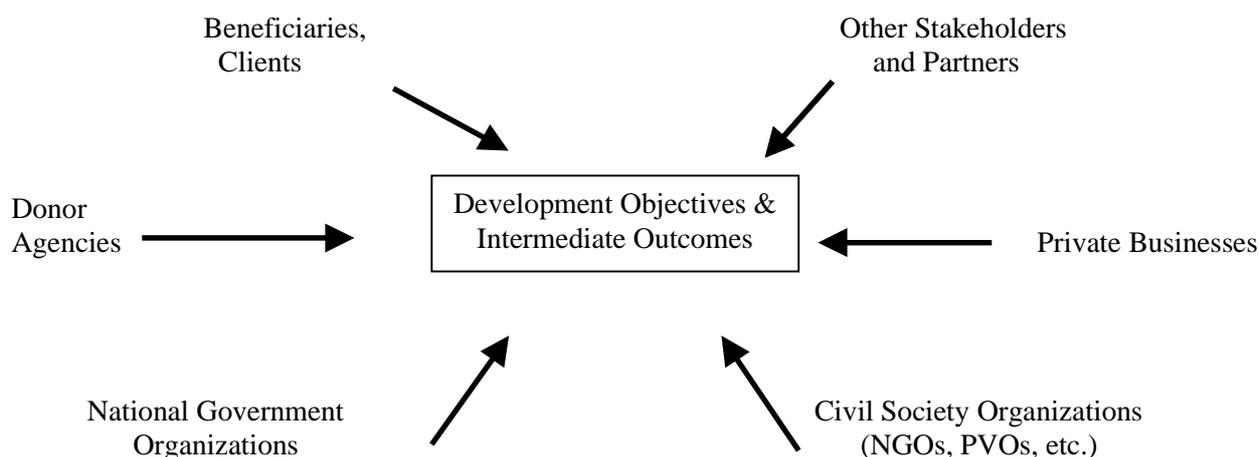
As management tools, results frameworks can be used by a donor's country operating units for:

- Reaching agreements (e.g., performance contracts) with donor agency headquarters concerning expected results and required resources for their achievement.

- Reaching consensus with other development partners in-country concerning divisions of responsibilities for achieving specific results (intermediate outcomes).

- Identifying and designing sets of related activities (project and non-project) to achieve intermediate outcomes for which they are responsible.

- Determining how results for which they are accountable will be monitored, selecting appropriate indicators, and developing the operating unit's overall performance monitoring system.

- Using the performance data to inform the operating unit's programming decisions (e.g., making corrective adjustments to specific activities, shifting resource allocations to more promising activities, etc.) and for reporting performance/results to agency headquarters.

The Process of Developing Results Frameworks: Developing a results framework works best if it is part of a collaborative or joint strategic planning process with those other development partners in the country working towards a shared development objective. Box 26 illustrates some of the key partners and actors that may influence development objectives and intermediate outcomes. Co-ordination or allocation of responsibilities among partners for achievement of specific intermediate outcomes can take place in this context. The developing country government should play a central, lead role in the process.

Box 26: Multiple Partners Influence Development Objectives and Intermediate Outcomes

Beneficiaries, Clients

Other Stakeholders and Partners

Donor Agencies

Development Objectives & Intermediate Outcomes

Private Businesses

National Government Organizations

Civil Society Organizations (NGOs, PVOs, etc.)

Broad ownership of a joint results framework (e.g., by partner government organizations, various donors and their implementing agencies, NGO and private sector actors) may promote greater co-ordination or harmonization of project and non-project activities all aimed at a shared development objective. Although this process takes time, a jointly developed results framework will be more comprehensive and representative with everyone's participation. However, the practice has often been less than the ideal. In fact, there may be a growing danger of competing, duplicative frameworks being pushed.[19] Perhaps this is because co-ordination tends to be so time-consuming as well as labor-intensive, or because donors tend to be most concerned with their own activities, or with promoting their own brand of framework. Also, the participatory approach calls for giving up some "control". Too often these frameworks have tended to be "donor-centric" -- that is, focused pretty much on the country program interests of a single donor.

Once a fully-joint results framework has been developed via a participatory process under the partner country's leadership, individual donor agency operating units may then want to prepared their "own" results frameworks as sub-sets of the overall framework. A unit's framework might then appropriately concentrate on those intermediate outcomes and outputs for which the donor unit has taken special responsibility, and would form the basis of their own program performance measurement system. That is fine, as long as it is a sub-set of (and thus clearly linked to and aligned with) a more comprehensive and jointly prepared country program results framework, and not one developed in isolation.

Box 27 provides some useful tips on how to develop a country program results framework.

Phase 2: Selecting indicators

Much of what has been said earlier about selecting indicators at the project level is equally valid at the country program level. So in this section, only some distinguishing characteristics will be noted.

Performance indicators for results frameworks have typically concentrated on selecting and measuring indicators at the higher-levels of results – that is, for development objectives and for intermediate outcomes. Less attention, perhaps unfortunately, has usually gone into linking specific project activities to the framework (and thus integrating indicators of project outputs, processes and inputs/expenditures within the broader framework). While this keeps the performance measurement system simpler and may allow for more flexibility at the project or activity level, it has drawbacks. For example, such a framework of high-level results may not be grounded in a realistic and well-conceived strategy that identifies the means to be used for achieving those results. It makes it difficult to relate achievement of objectives/results with needed resources and activities.

---

19.   See DFID, *Strategies for Sustainable Development: Can Country-level Strategic Planning Frameworks Converge to Achieve Sustainability and Eliminate Poverty?,* Background Briefing (September 2000)

Box 27: Developing Country Program Results Frameworks

Following these steps to developing a results framework may be useful:

Step 1: Clarify the development objective. Jointly developing and agreeing to a clear statement of the development objective will help partners gain a shared sense of ownership and responsibility for its achievement. In developing the objective statements, keep them results-oriented, unidimensional, precise and measurable.

Step 2: Identify the key intermediate outcomes. Next, identify all the intermediate outcomes or results that must first be accomplished in order to achieve the shared development objective. It is useful to clarify which partner organization(s) are going to be responsible for each intermediate outcome. (This will depend on individual agencies' areas of emphasis – i.e., their own organizational goals/priorities and areas of comparative advantage). This is the level of result (beyond outputs, of course) over which individual partners can exert reasonable influence, and for which it may be most appropriate to hold them accountable (i.e. the future performance of the responsible unit/partner will be judged based on its achievement). Developing outcome statements should follow the same criteria as with development objectives (above).

Step 3: Clarify the causal linkages between results. The next step is to clarify the principle causal connections between the intermediate outcomes and the strategic development objective. There may be multiple levels of intermediate outcomes as well as causal relationships between intermediate results at the same level.

Step 4: Identify critical assumptions. These are conditions outside the control or influence of the various partners (i.e., they are not results), which are likely to affect the achievement of results in the results framework (e.g., weather patterns, political stability, world prices, etc.)

Step 5: Finalize the Results Framework. Review the Framework to ensure it is not overly complicated. Re-check that objectives and outcomes are stated as results, are unidimensional, clear and precise, and are objectively verifiable/measurable. Assess if they are feasible given the anticipated resources available to the partners.

After completing the Results Framework, next steps involved in performance measurement include identifying indicators, establishing baselines and setting targets, developing performance monitoring plans, collecting and analyzing the data, etc. It is also time for the partners and donor operating units to design and implement the sets of activities and outputs for which they are taking responsibility.

Source: Adapted from USAID, *Building a Results Frameworks*, draft PM&E Tips.

Types of indicators

- *Indicators of development objectives* in the results framework are akin to project impacts in the project logframe, in that they are both generally conceptualized as long-term and significant, sector or sub-sector development results. Development objectives in results frameworks differ primarily in perspective -- they are viewed more explicitly as the consequence of multiple intermediate outcomes resulting from many different sets of project activities by multiple donors and partners, rather than from the perspective of one project.

- *Intermediate outcome indicators* in results frameworks are similar in concept to outcomes in the project logframe -- both are medium-term results that are more directly linked to specific partner/agency activities. They are both intermediate levels in the cause-effect chain, in between activity outputs on the one hand, and long-term development objectives on the other hand. However, results frameworks are far more comprehensive and explicitly cover all relevant intermediate outcomes that are necessary and sufficient to achieve the development objective. Thus there are many more intermediate outcomes with associated indicators that must be selected and ultimately measured (although the burden of monitoring them will be shared by different donor agencies and partner organizations).

- *Output indicators* in results frameworks (where they are explicitly included) are similar to output indicators in the project logframe, but are again much more comprehensive. They are inclusive of all outputs from all projects and non-project activities, grouped according to the intermediate outcomes to which they contribute. Developing and collecting data on output indicators are primarily the responsibility of the individual partners producing the outputs.

An ideal approach would be to include indicators for all hierarchy levels in one comprehensive results framework. By being more inclusive and balancing needs for indicators at all levels, one can avoid the mistake of focusing too exclusively at the higher-order impact level. The indicators at the very top of the hierarchy are too long-term and broad to be of much use for guiding shorter- or medium term project and non-project activity decisions, and moreover aren't much good for measuring individual donor contributions (attributions). In particular, more attention to developing good intermediate outcome indicators that are beyond outputs but still can be linked to individual donor activities/contributions, is needed. A number of levels of intermediate outcomes between outputs and ultimate impact may be needed to adequately demonstrate and measure the logical cause-and-effect chain.

The "ideal" results framework (and its related performance indicators) would be broad-based -- i.e., inclusive of the outputs and outcomes of all relevant donors' and other partners' activities. Responsibilities for collecting data for such a comprehensive number of performance measures would, of course, have to be divided among the involved donors/partners, or be co-sponsored, where appropriate, as discussed later.

While taking a comprehensive approach is the ideal, it may create practical difficulties, in terms of keeping the system reasonably simple. Also, it could require considerable time and effort for co-ordination among various partners working towards a shared development objective in a country. It requires joint indicator selection processes, particularly to agreeing to indicators and targets at the higher hierarchy levels of the results framework (as well as useful contextual indicators). In summary, it requires that partners are willing and able to share a similar approach (i.e., a jointly developed results framework, similar terminology, shared indicators, etc.).

Phase 3: Setting targets

Donor agency operating units using the country program approach to strategic planning often set targets (explicit values to be achieved within specific timeframes) for development objectives and intermediate outcomes in their results frameworks.

Setting targets at the level of country development objectives is best done jointly -- in other words, should be agreed to by all development partners working towards the same objective (under the leadership of the partner country government), drawing on national strategic planning targets, if available. These targets should take into account the likely effects of all donors' and other partners' activities and resources combined. Progress towards country development objectives/targets is the consequence of joint efforts, rarely if ever individual agency efforts. Accountability for achieving any targets at this level should thus also be viewed as a joint, not individual agency, responsibility.[20] A starting point for setting country-level sector targets might be to chose those already jointly agreed to by the development community at the global level under the strategy for *Shaping the 21st Century*. Country-specific targets, however, might have to be adjusted in most cases in light of country baselines (starting points) and expected rates of progress given country capacities, contexts and resource availability.

Targets at the intermediate outcome level, if primarily the expected result/consequence of a single donor's efforts, could be appropriately set by the donor operating unit that's responsible, although it is still best done in a participatory mode.

Setting targets at lower project-specific levels (e.g. outputs, processes, inputs) are the immediate responsibility of the individual donor operating unit involved (and its implementing agencies).

Phase 4: Monitoring Country Program Performance

To monitor performance within the comprehensive country program results framework approach just described, requires multiple data collection strategies to cover all hierarchy levels, and appropriate co-ordination and allocation of data collection responsibilities among partners. Such a comprehensive performance measurement system could become quite complex and requires considerable attention to co-ordination of data collection efforts. On the other hand, it offers opportunities for sharing the cost-burden of undertaking potentially expensive surveys, for co-ordinating and strengthening data collection capacity-building efforts within partner country statistical units, for developing more comparable measurement approaches, and for sharing performance data among donor and partner agencies.

Collecting data to measure change at the development objective (impact) level often requires conducting large-scale sample surveys periodically over time. Moreover, collecting data on the country context, important for assessing changes in assumptions and risk, may also require sample surveys. Such efforts can be expensive, time-consuming and technically complex. It often takes a long time before results at the sector impact level become evident, and may require a relatively long time-series of data to accurately assess trends. (For an example, see Box 28). However, if the whole strategic planning process -- i.e., development of the results framework and planning of the performance measurement system -- has been a joint process, opportunities for

---

20.   The practice in USAID of having operating units set their own targets for strategic objectives that are often well beyond their control, is now resulting in difficulties with oversight agencies. Not only is the operating units' performance being judged based on achievement of these targets, but also proof that results can be attributed to USAID programs is being demanded, which is very difficult. Setting targets apparently implies accountability for achieving them, from the auditors' perspectives.

co-ordinating such survey efforts and for cost-sharing should arise. Co-ordination among donors and partners with a stake in the impact data will mean reduced duplication of efforts and less burden on partner country statistical units. Moreover, co-ordinated support for survey efforts should contribute significantly to strengthening the partner country's statistical data collection capacities.

Box 28: Illustration of the Long Timeframe Needed to Collect Impact Data

Demonstrating change at the development objective level may take a long time and considerable effort. For example, Casley and Kumar demonstrated how effects and impacts of an agricultural/rural development program may take years before they become apparent and thus requires collection of time-series data over many years.

"Even a simple trend in crop production attributable to an intervention may take years before it can be measured. Casley and Kumar (1987, p. 118) constructed a table that shows the number of years of high-quality data required to determine a distinct trend with a given level of accuracy and statistical confidence. Their calculations demonstrate that in order to detect a rising trend in production of 4% per time point, with an accuracy of 25% either side for 95% confidence, twenty-one points are required – equivalent to twenty-one years of data for annual cropping."

This example of rain-fed agriculture is not unique. In a wide range of development projects, impact-level results may take many years to generate and are often influenced by unpredictable environments. Analysts may need impact data for numerous time periods to adequately assess performance trends.

Source: Krishna Kumar, "Measuring the Performance of Agricultural and Rural Development Programs", in *Evaluating Country Development Policies and Programs: New Approaches for a New Agenda,* AEA, 1995.

Experience with conducting large-scale surveys in developing country conditions has often been plagued with problems. Some of these difficulties are summarized in Box 29. The problems often encountered with larger surveys should not lead donors to conclude that such efforts will never work or aren't worth the effort. However, experience and lessons learned from the past can help to avoid past mistakes and improve future efforts. Also, experience should alert donors and other partners to the magnitude of the task and to its cost and time implications, and to the need for capacity-building training and technical expertise. It should also indicate the need and advisability of collecting more intermediate-level outcome data as well, which generally involves simpler, more rapid and low-cost methods. Often this interim data can be used as "leading" or proxy indicators for higher-order change.

<div style="border:1px solid black; padding:1em;">

Box 29: Problems Plaguing Large-Scale Surveys

A number of problems have persistently plagued large-scale surveys for development programs during the past few decades. Reviews of experience, summarized by Kumar, reveal the following problems were common:

1. In the absence of an analytical framework, variables were inappropriately identified and poorly conceptualized. In their enthusiasm for comprehensiveness, survey designers often weighed their design with too many independent, intervening and dependent variables, resulting in costly and unrealistic designs.

2. Little attention was given to making the concepts operational, and the issue of validity of indicator definitions was largely ignored

3. In many instances, sampling procedures were basically flawed. The sample was unnecessarily large, creating data collection and management problems.

4. The large sample size, the poor training of enumerators, and inadequate supervision resulted in high non-sampling errors, leading to dubious findings.

5. In most cases, surveys were not completed in time. Initial enthusiasm evaporated after the baseline survey, and follow-up surveys were not taken at all.

6. Even when data were gathered, they were not analyzed in time. Usually raw data were presented to the stakeholders without analysis. No attempts were made to translate the findings into concrete recommendations for actions.

7. Under these conditions, managers simply ignored the findings even when they were communicated.

Source: Krishna Kumar, "Measuring the Performance of Agricultural and Rural Development Programs", in *Evaluating Country Development Policies and Programs: New Approaches for a New Agenda,* AEA, 1995

</div>

Data on intermediate outcomes can often be collected in the shorter-term and using lower-cost and less complex methods, such as via mini surveys, customer surveys, community interviews, direct site observation, and other rapid appraisal techniques. Moreover, because intermediate outcome data are, in many cases, more directly and obviously connected to specific donor activities, they are preferable in terms of demonstrating the donor operating unit's individual performance and accountability. While there may be some opportunity for conducting some of these data collection efforts jointly, in most cases it will make most sense for donors/partners to divide up the data collection responsibilities for intermediate outcomes. In other words, each donor /partner would be responsible for collecting data on those intermediate outcomes for which they have taken the primary responsibility for producing (i.e., are relatively direct consequences of their own program outputs).

Data on outputs, to the extent that they are included in results frameworks, are collected routinely via individual project records and management systems, and are thus the obvious responsibility of individual donor agency operating units and their implementing agency partners.

Phase 5: Reviewing and reporting country program performance

Analysis, review and reporting of country program performance would be most effective if done in a participatory or joint fashion, involving all relevant donor operating units and partners from the developing country that share a development objective in a country. Thus far, most experience with analysis, review and reporting of country program performance has been from the narrower perspective of an individual donor agency's country program (Such as USAID's experience discussed below). However, several collaborative or joint efforts are now underway. For example, the DAC Working Party of Aid Evaluation has proposed that a series of country sector assessments should be conducted jointly by Members that might focus on key development goals articulated in the internationally agreed strategy for *Shaping the 21st Century*. To facilitate such joint assessments, Danida prepared a "Sector Matrix" for the DAC WP-EV that array Members' involvement in particular sectors and developing countries. In 1999, the first joint sector assessment was initiated that assesses the combined effects of donors' support to the transport sector in Ghana. Participants included Denmark, Germany, Japan, the Netherlands, the United Kingdom, the European Commission, the World Bank, and Ghana, which is chairing the Steering Group jointly with Denmark.

USAID has considerable experience with analyzing, reviewing and reporting performance at the country program level. USAID's operating units no longer report to headquarters on individual project performance. Instead, units submit annual *results review and resource request* (R4) reports that assess progress towards country-level strategic objectives. These R4 reports provide results data and indicate whether actual results surpassed, met or fell short of expectations, discuss actions being taken to address programs not meeting their targets, and contain the unit's request for funding needed to meet future targets. R4s feed into USAID's annual budget cycle. The intention is to ensure a closer link between program performance and resource allocations. Reviews of R4s have taken place annually at headquarters involving the operating units and regional bureaus. These reviews provide opportunities for joint review of performance, of needed corrective actions (or possibly further study/evaluations), and resource needs. However, their preparation plus the review process has become very labor-intensive. Data on actual versus expected results (targets) for strategic objectives and intermediate outcomes are extracted from the R4s and entered into a computer database for analysis and reporting in the agency's *annual performance report*.

These performance measurement systems at the country program level have taken time to establish. Over two-thirds of USAID's operating units now have actual performance data to compare and report against targets, up considerably from previous years.

A number of issues currently surround the R4 process. First, from the perspective of the reporting operating units (and also of their implementing partners who bear the brunt of data collection responsibilities), the R4 process has been overly complicated and time-consuming. A review of the process has recommended steps be taken to simplify the R4s – making them shorter, with fewer data and indicators, and to reduce formal annual USAID/Washington reviews to once every three years instead of annually.

Other issues have emerged from GAO reviews and Inspector General (IG) audits of the USAID operating units' R4s and of the programme performance monitoring systems on which they are based. The IG's key issue has been with the quality of performance indicators and data collected and reported; that is, concerns with accuracy, reliability, timeliness, etc. USAID has recently issued additional guidance on quality concerns, but

much additional work needs to be done by operating units if quality is to be improved.[21] USAID and its operating units also face external pressures to better demonstrate that the significant development results being reported (at the strategic objective level) are attributable to its activities. This is quite a challenge given that USAID units are often only minor actors affecting the observed development results. Demonstrating attribution is also complicated by the fact that USAID country program results frameworks only require monitoring performance at the highest results levels, often effectively de-linking them from activities. In other words, project logframes and related project-level monitoring are not consistently practiced and are not usually linked with the broader results frameworks. For example, a GAO report notes "USAID missions still have difficulty linking their activities to the broad indicators of development ... One way to provide a more complete picture of program performance could be to supplement performance measurement data with impact evaluation studies." [22]

These issues currently faced by USAID illustrate the potential for conflicting demands and pressures. On the one hand, pressures from operating units and implementing agencies/partners seek to make program performance monitoring, reviewing, and reporting systems simpler and geared towards management uses in the field. On the other hand, external sources such as the IG and GAO, who view the process more from an accountability reporting perspective, pressure for yet more data collection and analysis efforts to ensure better data quality and demonstration of USAID's contribution.

Analyzing program attribution

Monitoring performance data trends at the intermediate outcome and development objective levels alone cannot demonstrate attribution to a single donor's program activities because of extraneous factors. Especially at the development objective level, the interventions of other donors and partners are likely to be major influences on the results as well. Attributing results to a particular program (i.e., measuring the net impacts) has proven to be a difficult analytical challenge.[23] Box 30 addresses the analytical difficulties encountered by the World Bank in attempting to measure net impacts of agricultural and rural development (ARD) programs, and illustrates the nature of these difficulties more generally. Use of methodologies such as quasi-experimental designs and multiple regression analysis to control for exogenous factors have in many cases failed. Conducting complementary in-depth evaluations can help demonstrate attribution, by re-examining the program's underlying logic and assumptions, and by providing evidence factoring out extraneous influences. For judging performance of individual donor programs, use of output data or lower-level intermediate outcomes with fairly obvious links to an agency's programs might be more suitable and less susceptible to the attribution problem.

---

21.   See USAID, "Guidelines for Indicator and Data Quality", *Performance Monitoring and Evaluation Tips*, Number 12, 1998.

22.   U.S. GAO, *USAID Reengineering at Overseas Missions* (1997).

23.   If possible, it would be best to convince oversight agencies and auditors that it is not appropriate to judge an individual donor operating unit's performance at the country development objective level, thus avoiding the analytical problem of demonstrating attribution at this high a level.

Box 30: Measuring Net Effects and Impacts of ARD Programs

"A critical, though not always explicit, assumption behind the PMS (performance measurement system) model is that a set of performance indicators can measure the results attributable to a program. Although this assumption may be correct when the focus is on outputs, it becomes doubtful when the results are defined in terms of higher-order effects and impacts of programs. Only in a few cases can performance indicators alone assess the effects and impacts that have directly resulted from ARD (agricultural and rural development) programs and not from extraneous factors.

The textbook approach for measuring effects and impacts is to follow either of two methodological strategies: the quasi-experimental design or controlling the effects of exogenous factors by statistical tools, particularly multiple regression analysis. Both approaches use a set of indicators, but extensive reviews of ARD projects show that neither approach has succeeded. ARD interventions have invariably failed to measure the "net impacts" of a project, much less a more complex program. Even when serious attempts have been made, they have been based on heroic assumptions and questionable data.

Determining whether a program is producing the intended effects and impacts requires the following:

- Time-series data on selected indicators. For example, a national agricultural extension program designed to increase production requires production estimates for the geographical areas covered by the program.

- A periodic re-examination of the key assumptions of the underlying intervention models. Re-examination is necessary because the settings and contents of programs evolve over time. The evaluators of an extension program, for example, must routinely re-examine assumptions about the efficacy of the technical packages and the delivery system to determine whether the assumptions are still justified.

- Some evidence, not necessarily quantitative, to factor out the possibility of extraneous variables having largely produced the observed changes. Evaluators of the extension program should be able to establish from key informant interviews, previous studies, and direct observation methods whether or not the observed change in cereal production was largely attributable to non-program factors, such as improved marketing system.

A carefully designed and judiciously implemented PMS model can in most cases generate time-series data on selected indicators, but it cannot adequately respond to the remaining two requirements, which usually require supplementary analyses. For those, evaluators may need to conduct in-depth evaluation studies to complement performance indicators in assessing attribution or in evaluating the continuing relevance of program outputs, targets and objectives in a changing environment."

Source: Quote from Krishna Kumar, *Measuring the Performance of ARD Programs*, op cit, 1995.

This analytical challenge is not unique to the development agencies, although the nature of development work may compound or magnify the challenge. For example, in a recent review by the U.S. General Accounting Office (GAO) of U.S. federal government agencies, separating the net impact of a program from other factors external to a program was identified as the number one challenge when it came to analyzing and reporting performance.[24] U.S. federal agencies were using a variety of techniques in responding to this challenge, including the following approaches to analysis:

- Specifying as outcomes only measures that the program can directly affect

- Advising the use of control groups

- Using customer satisfaction measures

- Monitoring results at the regional level

- Expanding data collection to include potential outcome variables

- Analyzing time-series data

- Analyzing local level effects that are more easily understood

- Involving stakeholders

Notably, these techniques were often employed at the sub-national level where the influence of other (non-program) variables was either reduced or easier to observe and control for. For example, using disaggregated, regional data enabled analysts to better determine the cause-effect relationships between programs and its effects. Others sponsored special studies in selected localities to identify a program's impact at the local level, where it can control for more factors. Others minimized the influence of external factors on their program's outcomes by selecting performance indicators that were "quite proximate to program outputs, permitting a more direct causal link to be drawn between program activities and results". Yet others used measures of customer satisfaction.

Aligning Project Logframes with Results Frameworks

How does a donor operating unit move from a having a portfolio of unrelated, stand-alone projects to an approach that emphasizes well-defined programs drawing together numerous project activities to address significant development objectives? First, they will need to integrate their existing projects into the new country program results framework, as best as possible. This may not necessarily be a simple or straightforward task, and may have to be done over time in an iterative manner. This is a positive process, however, leading to a much more focused or strategic approach to programming. Old project activities that do not contribute to the few new country-level development objectives to which the operating unit is now committed will either eventually die out, be terminated, or be redesigned to focus more clearly on these priority objectives. This is part of a "focus and concentrate" phase during which project resources and activities are increasingly aligned with the strategic development objectives. The unit's focus or vision is re-directed from implementation to higher-order results, and from stand-alone projects to coherent programs.

In this process, however, it would be a mistake to forget completely about project logframes and project-level design and implementation concerns. By focusing on higher-order outcomes and impacts, the results

---

24.   U.S. GAO, *Managing for Results: Analytical Challenges in Measuring Performance*, GAO/HEHS/GGD-97-138, May 1997.

frameworks do not have the project-specific details about costs, other inputs, activities/ processes, and even sometimes outputs that are available in project logframes -- details that are useful for guiding project implementation.[25] What is useful, however, is to use both tools and ensure an *explicit* connection or integration of project logframes into the broader results frameworks. This will make clear to everyone (including the project implementing agency and other partners as well as operating unit staff) how a particular project is contributing to the broader country program strategy and overall development objective. Typically -- but not always -- the project goal will be the equivalent of the development objective, while project purposes will equate with intermediate outcomes.[26] Box 31 provides an illustration of how a project's logframe (marked in bold) might be aligned within a broader country program results framework.

Shortcomings of the country program results framework approach

While there are many potential advantages to using a country program results framework approach to performance measurement and management, there are also possible drawbacks, if applied too narrowly. For example, in practice results frameworks have often focused too exclusively at the highest-levels of results. This may result in performance measurement systems that monitor results over which a donor agency operating unit has little practical control and thus cannot realistically be held accountable for achieving. Moreover, even if it were appropriate, data collection and analytical problems make demonstrating attribution very difficult at the impact level. Concentrating on measuring results at the impact level opens a Pandora's box of accountability and attribution problems for donor agencies and their operating units. Also, because impact level data typically takes years to register improvements, it may not be very useful for management's shorter-term decisions or for their needs to demonstrate and report on progress more frequently, such as in annual performance reports.

Data on project outputs in some cases (e.g., in USAID) may not even be collected or reported to headquarters, so there is no record of performance at this level of results for which the agency is fully responsible and accountable. Results data at the intermediate outcome level is also often weak. Moreover, focusing too exclusively on the higher levels of objectives may lead to a "disconnect" between high-level results on the one hand, and means (resources, activities) on the other. Explicitly linking project logframes with results frameworks may not be happening regularly. Too much attention on monitoring performance at the top may mean monitoring of implementation performance is being ignored.

---

25.   While useful to project managers, such detailed project/activity implementation monitoring information need not necessarily be reported to headquarters.

26.   There may be a mis-match between hierarchy levels of the project logframe and the results framework; it may be necessary to go back and re-design existing project logframes to better conform to or align with the new country results framework.

Box 31: Illustration of How a Project Logframe Might Align Within a Country Program Results Framework

(Project Goal)

Development Objective
_____

Stem the Loss of
Biodiversity

(Project Purpose)

Intermediate Outcome	Intermediate Outcome	Intermediate Outcome	Intermediate Outcome
Encroachment of Protected Areas Reduced	Approaches to generating income and employment from non destructive resources developed	Capacity of local and non-governmental organizations to plan and manage natural resources strengthened	Policies to encourage resource management (rather than destruction) adopted and implemented

(Project Outputs)

Outputs	Outputs	Outputs	Outputs
Resource monitoring	Park boundaries	Park personnel trained	Park administration and

Another potential problem is the tendency of agency operating units to develop their country program results framework in relative isolation and from their own agency's perspective alone. While there may be some limited participation from their implementing agencies and other stakeholders, the focus is usually on their own agency's programs and strategies, rather than placing equal focus on all relevant partners' programs that contribute to a given development objective. Country-level results frameworks can best be developed in a highly collaborative or joint fashion with other key donors and partners, under the leadership of the partner country government. Later, it would be appropriate for individual donor operating units to prepare their own results frameworks and program strategies as sub-sets of the larger joint framework, concentrating on those intermediate results for which they are going to be responsible for achieving and monitoring results.

Such a scenario of collaboration, however, assumes that donor agencies and partners are all prepared to use the same methodological approach and terminology, and are willing to participate in and share control over the process of developing a jointly owned framework. This may be complicated by the fact that some donors have not yet adopted a country program results framework approach, while others appear to be advocating their own, possibly competing versions. Some agencies have already "invested" in their own approaches, frameworks, terminology, common indicators, etc. and thus harmonizing these efforts may not be that easy. Moreover, partner country organizations may not be familiar with or committed to performance measurement and management approaches and tools. Building their capacity for performance measurement and management may be a pre-requisite before they can "take over the driver's seat" in leading and co-ordinating country program strategic planning and performance monitoring systems. Other constraining factors include the potentially time-consuming and labor-intensive nature of collaborative processes. Finally, government-wide approaches to performance measurement and reporting within the OECD countries may also constrain the flexibility that donor agencies have to co-ordinate with others at the partner country level.

In conclusion, the ideal is to develop a comprehensive, balanced, and jointly-prepared country program results framework that is inclusive of all hierarchy levels and also of all relevant partners' activities who share a particular country development objective. It may be difficult to achieve, but worth striving for. Comprehensiveness may need to be tempered to some extent in the interests of keeping the system relatively simple and feasible. Once a comprehensive, joint framework is collaboratively prepared, individual donor units could then prepare their own results frameworks as essentially sub-sets of the larger joint framework. The data collection burden, as well as the program implementation responsibilities themselves, then could be allocated among donors and other partners, or shared via joint efforts where that makes most sense. Unless donors learn to better harmonize these efforts, there is a real danger that as they shift to a country program approach, their competing frameworks, overlapping indicator systems, diverse terminology, and duplicative demands for results data may actually over-burden partner countries rather than assist them.

V. PERFORMANCE MEASUREMENT IN THE DEVELOPMENT CO-OPERATION AGENCIES

-- *The Agency Level* --

The establishment of strategic planning and performance measurement systems at the corporate (agency-wide) level has recently become an urgent priority in all of the donor agencies reviewed. This has been driven by growing public pressures calling for government accountability, and especially for transparent reporting of results achieved by government agencies in the OECD countries. In most cases, the reforms are mandated by recent government-wide legislation or executive orders requiring results based management approaches. The donor agencies are responding by clarifying their overarching goals and seeking ways to summarize their achievements vis-à-vis those goals.

Obstacles to performance measurement at the corporate level

Measuring and reporting on results at the agency-wide level poses a significant challenge for the donor agencies. They face a number of obstacles in attempting to aggregate results, that are in some respects different from other government agencies, complicated by the nature of development co-operation work. (See Box 32 for a discussion of the special challenges facing donor agency efforts to aggregate results.)

This chapter explores how the donor agencies are attempting to respond to the challenge. Their approaches are quite varied -- there's much more diversity at the corporate level than at the project level (where most agencies have followed similar approaches based on the logframe) or at the country program level (where few agencies have yet ventured). Because approaches for measuring corporate-level performance vary so much from agency to agency, illustrations are provided frequently in this chapter, comparing or contrasting approaches taken by individual donor agencies.

Phase 1: Formulating objectives

All of the donor agencies reviewed have taken steps recently to clarify their corporate level goals or objectives -- so they can more clearly articulate to the taxpaying public, legislative bodies, and oversight agencies the overall aims of their development assistance efforts. These statements about agency goals and objectives have also served to provide a framework or structure for reporting on overall agency results being achieved. This is viewed as important in an era of declining aid budgets, increasing competition for funding, and growing public scepticism about the effectiveness of development aid. Clarifying agency-level goals has also been useful as an internal management tool for strategic planning – that is, for focusing and concentrating the agencies' assistance portfolio and resources within its priority goal areas. Country operating units (i.e., country field missions, posts or offices) in many cases have been asked to align their country program objectives and/or their projects within the new corporate goal structures.

Box 32: Challenges Facing Donor Agencies' Efforts to Aggregate Results

Measuring and aggregating results across country programs and projects in order to assess and report on performance at an agency-wide level is challenging for the donor agencies for a number of reasons, many of which are inherent in the nature of international development co-operation work. For example:

Unlike most domestic government agencies, the donor agencies operate not in one country but in many. This means there is no single national source of comparable measures and data for any given substantive area. While a number of international organizations have established international databases that attempt to provide standard indicator data across countries, their comparability remains in most cases weak.

Moreover, the poverty conditions in the partner countries where development agencies work mean typically inadequate data collection infrastructures and capacities with consequent problems in data quality, coverage and timeliness.

Most government agencies' experiences with performance measurement has been with direct service delivery types of programs, whereas development co-operation agencies are increasingly concentrating on institutional capacity development and policy reform rather than service delivery. Finding appropriate performance indicators for these types of programs remain a challenge.

Also, donor agencies are adopting new goals and program areas, such as democracy and good governance, where development theory, effective program strategies and good performance indicators have yet to be developed, agreed upon, and standardized.

Donor agencies are typically only minor actors in achieving higher-order outcomes and impacts. Partner country government agencies, NGOs, private sector businesses, and other actors may all contribute to achieving the same development objective. Thus, the practice of aggregating results based on changes in developing country national statistics and attributing them to the activities or "performance" of a single development agency is highly questionable.

The substantive range of sectors and activities in which the donor agencies are involved is typically very broad, covering education, health, population, agriculture, infrastructure, the environment, governance, poverty alleviation, etc., whereas most domestic agencies deal with a single sector and a narrower range of related activities. Consequently, establishing effective performance measurement systems to cover all these sectors, as well as numerous countries, becomes a very complex, time consuming, and expensive proposition for donor agencies.

Achieving development objectives is a long-term business, and it is generally not feasible to monitor the long-term impacts of current activities for quite some time. In other words, there is a lag factor, and thus annual performance monitoring of development results may not be very appropriate (whereas ex post evaluation is). Moreover, the expense of impact-level data collection makes annual surveys prohibitive or infeasible, as well as not very appropriate.

For those donor agencies that have decentralized structures with authority and responsibility for setting objectives and developing projects delegated to country operating units, there may be little comparability among project activities and their results indicators, with consequent difficulties in aggregating results across those activities. Ironically, the push for corporate-level reporting on results appears to be in conflict with the results based management concept of delegating authority and responsibility in the field.

Agency strategic frameworks

All of the donor agencies reviewed have recently developed and issued policy papers or strategic plans that clearly articulate the agency's overall mission and the key development goals or priority areas on which they will concentrate. Many of the agencies' goals are sector-oriented (e.g., better education, improved health, good governance) although some may be cross-cutting special concerns (e.g., gender equality, partnerships) or internal management efficiency goals. The number of key goal areas vary from agency to agency, ranging anywhere from 4 to 10. See Box 33 for a summary listing of the key goal areas of each of the donor agencies reviewed for which information was available. In addition to informing external audiences about agency goals and guiding internal management efforts to focus the portfolio on priorities, these high-level statements of an agency's goals are being used as frameworks or structures around which agency-wide performance is being measured and reported. These corporate frameworks have various names from agency to agency, for example:

- USAID's strategic framework.

- UNDP's strategic results framework.

- DFID's output and performance analysis.

- AusAID's performance information framework.

- World Bank's scorecard.

- Danida's output and outcome indicator system.

Annex 2 outlines each of these donor agencies' corporate-level strategic frameworks, including their agency goals and associated performance indicators and targets (where available). Below, an effort is made to synthesize these diverse agency experiences and approaches, pointing out their similarities and differences.

In some cases (e.g., USAID, UNDP, Danida), donor agencies have elaborated their key corporate goals into several sub-categories, forming a multi-level framework or hierarchy of objectives. These multi-level frameworks serve to clarify even further what an agency seeks to achieve and how it intends to contribute. The hierarchies serve as detailed structures for reporting on agency results at several levels.

The hierarchy levels have different names and definitions from agency to agency, and not all agencies have the same number of levels. Typical levels, which are referred to in this paper as agency mission, goals, sub-goals, and program approaches, might be generally described as follows:

- *Agency mission* - the overall purpose or aim of the agency's development assistance program.

- *Agency goals* - the priority long-term goals of the agency's development assistance program that support the overall mission.

- *Agency sub-goals* - medium-term agency objectives that contribute to agency goals.

- *Agency program approaches* - the typical program approaches or strategies that the agency's country operating units support in order to contribute to agency sub-goals.

Box 33: Comparison of Donor Agencies' Key Goal Areas

DFID's Output and Performance Analysis	USAID's Strategic Framework	AusAID's Performance Information Framework
Objectives: Policies and actions which promote sustainable livelihoods Better education, health, and opportunities for poor people Protection and better management of the natural and physical environment Departmental operations	**Goals:** Broad-based economic growth and agricultural development encouraged Democracy and good governance strengthened Human capacity built through education and training World population stabilised and human health protected World's environment protected for long-term sustainability Lives saved, suffering associated with natural and man-made disasters reduced, and conditions for political and economic development re-established USAID remains a premier bilateral agency	**Key Results Areas:** Improve agricultural and regional development Increase access and quality of education Promote effective governance Improve health of people Provide essential infrastructure for people Deliver humanitarian and emergency assistance Promote environmental sustainability Promote equal opportunities for men and women as participants and beneficiaries of development

Box 33: Comparison of Donor Agencies' Key Goal Areas (continued)

Danida's Output and Outcome Indicator System	UNDP's Strategic Results Framework	World Bank's Scorecard
Assistance Sectors: Agriculture Education Environment Good governance Health Infrastructure (transport, electrification, telephones) Water (water resources, drinking water, sanitation)	**Goals:** Promote decentralisation that supports participatory local governance, strengthens local organisations, and empowers communities Promote poverty focused development Equal participation and gender equality concerns in governance and economic and political decision-making at all levels Promote integration of sound environmental management with national development policies and programs Special development situations (crisis countries) UNDP support to the UN Management	**Development Outcomes (Tier 1.a):** Poverty reduction Equitable income growth Human development Environmental sustainability **Intermediate Outcomes (Tier 1.B):** Policy reform Institutional capacity Resource mobilisation **Strategy Effectiveness (Tier 2)** **Process and Capacity (Tier 3)**

Box 34 illustrates a typical structure of a multi-level agency strategic framework and the relationships among hierarchy levels. A few agencies (e.g., USAID) have found it useful to present their strategic frameworks as graphic diagrams (visual displays), using objective tree concepts. Box 35 gives an example for the environmental goal from the USAID strategic framework.

More detail on the typical hierarchy levels is provided below.

Agency mission (also called overall objective, statement of purpose, etc) -- states the overall purpose of the agency's assistance program. Usually it relates to promoting sustainable development or reducing poverty in the developing countries, which is considered to advance the donor country's national interest or foreign policy objectives.

Agency goals (also called key results areas, agency objectives, etc.) - represent long-term, broad development objectives for developing countries, generally relating to key sectors of assistance (e.g., improving health conditions, promoting good governance, improving educational opportunities) and cross-cutting objectives (e.g., promoting gender equality, supporting partnerships), but may also involve internal agency management goals. The goals directly support the agency's mission statement. Goals may reflect the agency's authorizing and appropriating legislation, administration priorities, consultations with key stakeholders, and/or the growing consensus among donors. Many agencies have adopted development goals that are the same or similar to those agreed to by the broader development community at various international conventions, synthesized in the DAC report, *Shaping the 21st Century: The Role of Development Co-operation* (May, 1996). See Box 36 for a summary of these internationally agreed goals, targets and indicators. Consequently, there is considerable similarity in goals across the donor agencies.

Agency sub-goals (also called agency objectives) - represent the next level of objective; they are medium-term objectives that contribute to achievement of the goal. There are usually several sub-goals under each goal. They tend to be at the sub-sector level, and more specific to the individual donor agency's priority areas of concentration -- e.g., organizationally mandated or legislated objectives, or areas of comparative advantage. For example, for a goal of promoting good governance, sub-goals might include supporting the rule of law, building civil society, supporting electoral processes, and reforming the public sector.
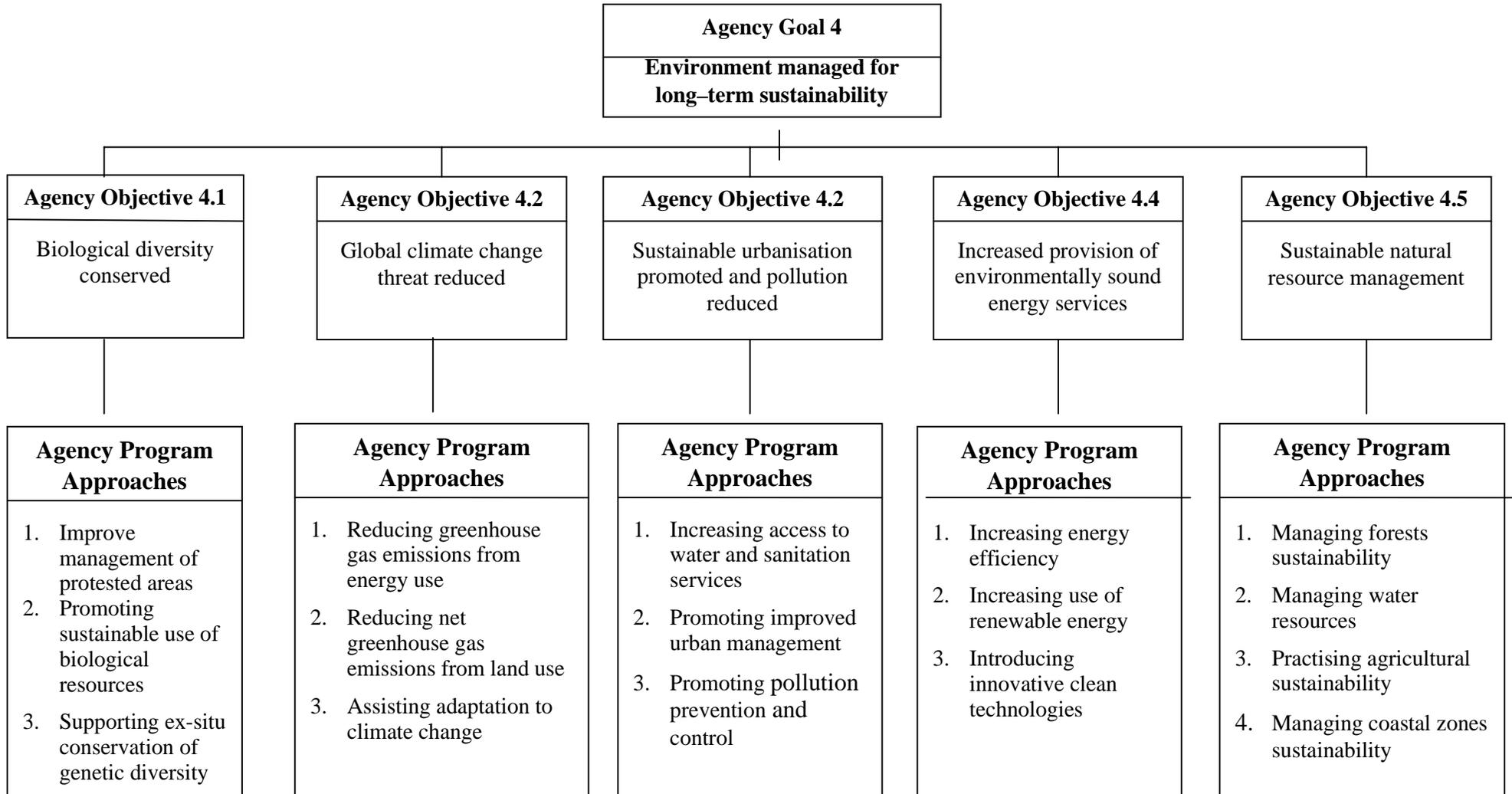
Agency program approaches (also called strategic areas of support) -- represent the typical program strategies or approaches used by an agency's country operating units for achieving agency sub-goals. Approaches usually pertain to specific categories of similar types of projects -- that is, an agency's commonly supported intervention approaches.[27] For example, for a sub-goal of improving the rule of law, program approaches might include improved management of court cases, improved working facilities for the judiciary, rehabilitation of court buildings, provision of legal aid, etc. Some agencies (e.g., Danida, UNDP) ask their country offices to report to headquarters on their project/program results (outputs and outcomes) categorized by these approaches, because at this level of detail, outputs and outcomes are generally similar enough to be comparable, thus facilitating aggregation.

---

27.   In some cases these approaches might be mandated or favored by headquarters, or they may simply reflect those approaches the field units are already supporting.

Box 34: Agency Strategic Framework Hierarchy Levels

```
        ┌─────────────────────┐
        │  National Foreign   │
        │  Policy Interests   │
        └─────────────────────┘
                   ↑
        ┌─────────────────────┐
        │       Agency        │
        │      Mission        │
        └─────────────────────┘
                   ↑
        ┌─────────────────────┐
        │       Agency        │
        │       Goals         │
        └─────────────────────┘
                   ↑
        ┌─────────────────────┐
        │       Agency        │
        │     Sub-Goals       │
        └─────────────────────┘
                   ↑
        ┌─────────────────────┐
        │   Agency Program    │
        │     Approaches      │
        └─────────────────────┘
```
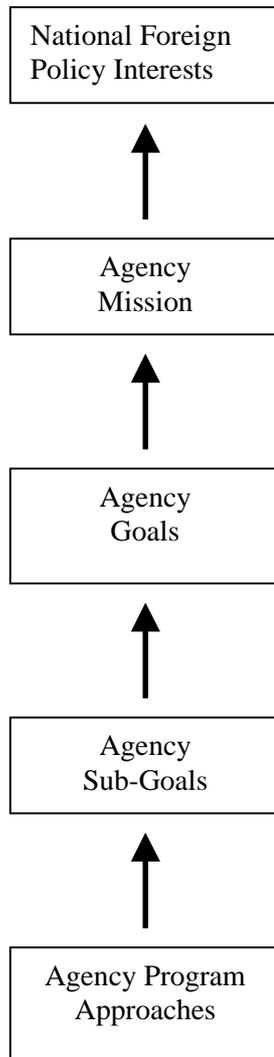
Box 35: USAID's Strategic Framework for the Environmental Goal

Agency Goal 4
Environment managed for long–term sustainability

Agency Objective 4.1	**Agency Objective 4.2**	**Agency Objective 4.2**	**Agency Objective 4.4**	**Agency Objective 4.5**
Biological diversity conserved	Global climate change threat reduced	Sustainable urbanisation promoted and pollution reduced	Increased provision of environmentally sound energy services	Sustainable natural resource management

Agency Program Approaches	**Agency Program Approaches**	**Agency Program Approaches**	**Agency Program Approaches**	**Agency Program Approaches**
1. Improve management of protested areas 2. Promoting sustainable use of biological resources 3. Supporting ex-situ conservation of genetic diversity	1. Reducing greenhouse gas emissions from energy use 2. Reducing net greenhouse gas emissions from land use 3. Assisting adaptation to climate change	1. Increasing access to water and sanitation services 2. Promoting improved urban management 3. Promoting pollution prevention and control	1. Increasing energy efficiency 2. Increasing use of renewable energy 3. Introducing innovative clean technologies	1. Managing forests sustainability 2. Managing water resources 3. Practising agricultural sustainability 4. Managing coastal zones sustainability

Box 36: Internationally Agreed Development Goals, Targets and Indicators

Goals and Targets	Core Set of Indicators
Economic well-being **Reducing extreme poverty** The proportion of people living in extreme poverty in developing countries should be reduced by at least one-half by 2015	• Incidence of extreme poverty • Poverty gap ratio • Poorest fifth's share of national consumption • Child malnutrition
Social development **Universal primary education** There should be universal primary education in all countries by 2015	• Enrolment in primary education • Completion of primary education • Adult literacy rate
Gender equality Progress towards gender equality and the empowerment of women should be demonstrated by eliminating gender disparity in primary and secondary education by 2005	• Gender equality in education • Gender equality in adult literacy
Infant and child mortality The death rates for infants and children under the age of five years should be reduced in each developing country by two-thirds the 1990 level by 2015	• Infant mortality rate • Child mortality rate
Maternal mortality The rate of maternal mortality should be reduced by three-fourths between 1990 and 2015	• Maternal mortality ratio • Births attended by skilled health personnel
Reproductive health and population Access should be available through the primary health care system to reproductive health services for all individuals of appropriate ages, no later than the year 2015	• Contraceptive prevalence rate • HIV prevalence rate
Environmental sustainability and regeneration **Environment** There should be a current national strategy for sustainable development, in the process of implementation, in every country by 2005, so as to ensure that current trends in the loss of environmental resources are effectively reversed at both global and national levels by 2015	• Countries with national environmental plans • Access to safe water • Intensity of fresh water use • Biodiversity: land area protected • Energy efficiency • Carbon dioxide emissions

Source: OECD/DAC, *Measuring Development Progress: A Working Set of Core Indicators*.

Agency strategic frameworks are useful management tools that have been used for the following purposes:

- To communicate to stakeholders and partners the essence of the agency's strategic priorities.

- To provide strategic direction -- for focusing and concentrating country operating units' assistance efforts (projects/ programs) on these agency priorities.

- To serve as a structure for categorizing, measuring and reporting on development results at the corporate (agency-wide) level.

- To contribute to strategic management decisions -- e.g., resource allocations among countries and sectors based on need and performance criteria.

Process of developing agency strategic frameworks

The process of formulating corporate-level goals is necessarily a political process, involving the agency's senior management and participation of key domestic stakeholders such as representatives from legislatures and senior staff from executive oversight and audit agencies. There may be legislation mandating specific goals for foreign assistance as well as legislation outlining requirements for government-wide strategic planning, performance measurement and reporting that must be followed. Some guiding principles to consider in the process of selecting and stating goals and sub-goals for an agency strategic framework are given below. To the extent possible, the objective statements should be:

- Statements of results, not means or actions.

- Precise and simple statements that can be easily understood.

- Measurable (objectively verifiable).

- Unidimensional (consisting of only one result per objective).

- Logically consistent among levels (reflect cause-effect linkages).

Following these criteria may not always be possible, since political considerations (i.e., what the agency's leadership wants to emphasize) may conflict with technical criteria (e.g., logically consistent, unidimensional, etc.). The over-riding concern should be to make the framework and its statement of objectives simple, clear and precise, and something that the agency's senior management will identify with, embrace as their own, and find useful for making strategic policy and programming decisions and for corpoate reporting on results to external stakeholder groups.

Phase 2. Selecting indicators

The corporate-level frameworks provide the structure around which donor agencies intend to measure and report on agency-wide results. The first step towards measurement involves the selection of indicators appropriate for this level. Generally, the same criteria that apply to selecting indicators at the project and country program levels apply at this level as well. However, a key factor in selection of indicators for the agency level is that they be comparable or standardized so that they may be summed up (aggregated) across projects or programs to higher organizational levels (e.g., for geographic departments, for key sector

goal areas, or for the overall agency portfolio).[28] Other important criteria for indicator selection are easy availability or accessibility of the data from existing sources.

Types of indicators

The key types of indicators that agencies have come up with so far for measuring corporate performance and results within their strategic framework structures are summarized below. They come from two basic sources -- widely available data from international statistical databases regarding country-level, sector development trends, and from their country operating units' performance monitoring systems regarding the performance and results of their projects or programs (e.g., outputs, outcomes, performance ratings).[29]

Key types of indicators used by the donor agencies for aggregating and reporting on agency-level performance and results are listed below. In many cases, the indicator data are aggregated or summarized at the level of agency goals or sub-goals. However, when attempting to add up project or program results (e.g., outputs and outcomes) across the agency, usually a more detailed hierarchy level is required. By categorizing at the level of program approaches, measures of outputs and outcomes may be similar enough to be aggregated.

Indicators aggregated at the level of agency goals (or sub-goals):

- Aggregation of country impact-level development statistics related to each of the agency's goals. Often they are the same or similar to the indicators agreed to by the international development community, reflected in the DAC report, *Shaping the 21st Century* (e.g., student enrolment rates, under-5 child mortality rates, percent of the population living in extreme poverty, etc.).

- Aggregation of project/program performance ratings within each goal area. (e.g., xx % of these projects/programs receiving a satisfactory or higher rating).

- Aggregation of project/program expenditures within each goal area (e.g., $xx million spent on these projects/programs).

- Aggregation of number of projects/programs implemented within each goal area.

- Aggregation of number of countries receiving project/program assistance within each goal area.

Indicators aggregated at the level of program approaches:

- Aggregations of project/program outputs achieved, categorized by agency program approaches. For example:

⇒ Number and types of goods and services provided (e.g., water supply connections, immunizations, health centers, training courses, etc.)

---

28. Unfortunately, this criteria may directly conflict with principles supporting participatory processes for indicator selection at the project and country program levels, which may result in non-comparable indicators.

29. In most agencies, the aggregation is based directly on project-level performance information. In agencies such as USAID and UNDP that have moved to country program-level results frameworks and performance measurement systems, the unit of analysis from which aggregations are made is the program, not the project.

⇒ Number of people assisted (e.g., numbers receiving health services, training, micro-finance loans, etc)

- Aggregations of project/program outcomes achieved, categorized by program approaches. For example:

⇒ Percentage of beneficiaries (clients) fully satisfied with goods or services provided (e.g., with training, schools, contraceptive supplies, basic health services, etc.) [30]

⇒ Short-to-medium term effects on beneficiaries (clients) of using the goods or services (e.g., increased yields of farmers practicing new agricultural techniques, improved test scores of children attending project schools, etc.).

These different types of corporate-level performance indicators each have their pros and cons. Most agencies have sought to balance these tradeoffs by selecting a mix of them. Some of the indicators noted above are really not performance or results-oriented indicators. Rather, they are indicators of the magnitude of effort going into achievement of a goal -- for example, the total expenditure on projects, the total number of projects, or the total number of countries with projects within a given goal area.

Some of the characteristics, including advantages and disadvantages, of different indicator options are discussed below.

1.   *Project/program performance ratings.* Project or program performance ratings are useful measures that have the important advantage of being able to aggregate or sum up across the agency's portfolio. Many agencies are using this approach to report on agency performance across the whole portfolio, within different goal/sub-goal areas, or within different geographic regions. There are, however, some problems with the meaning and comparability of ratings across a wide diversity of projects and country settings. Some agencies report problems with coverage -- i.e., not all projects are routinely being rated. Moreover, obtaining objective ratings based on self-assessments may be problematic, especially if managers fear reprisals or funding cutbacks for poor performance ratings. Some sort of validation process may be needed. For example, the World Bank has such a validation process. The OED reviews all self-assessments of projects (implementation completion reports) and adjusts performance ratings if necessary. Moreover, the OED conducts independent field-based audits for 25% of all completed projects.

Getting ratings on an annual (or more frequent) basis to monitor and report on progress of on-going projects (rather than just at project completion), means that results are often limited to achievement of project outputs or to prospective assessments of *likely* outcomes. Moreover there may be some question about how satisfying performance ratings are to the general public or to other external stakeholder audiences that the agencies are hoping to impress. These audiences may rather hear about concrete or substantive welfare improvements made, such as reductions in child mortality, increases in school enrolments, etc., rather than subjective performance ratings.

2.   *Project/program output and outcome indicators.* Reporting on the actual outputs and outcomes of projects or programs categorzed by goals, sub-goals, or program approaches is also practiced by some agencies. However, here the problem is usually how to aggregate across a great diversity of indicators. In some agencies (e.g., Danida) country units are asked to report results according to standard output indicators and (to a much lesser extent) standard outcome indicators that have been defined by headquarters. These standard results indicators are specific to individual categories of program approaches.

---

30.   None of the donor agencies appear to be systematically using this intermediate outcome indicator, although it would have the advantage of being able to aggregate across different types of projects/programs, and serve as a leading indicator for higher-order outcomes. It would require systematic conduct of follow-on customer surveys using a standard format.

(See Danida examples in Box 9.) Other agencies (e.g., USAID) have chosen not to require standard indicators that country operating units must report on, but instead offer guidelines recommending or suggesting use of standard indicators, if considered appropriate to their situation by the country operating unit.
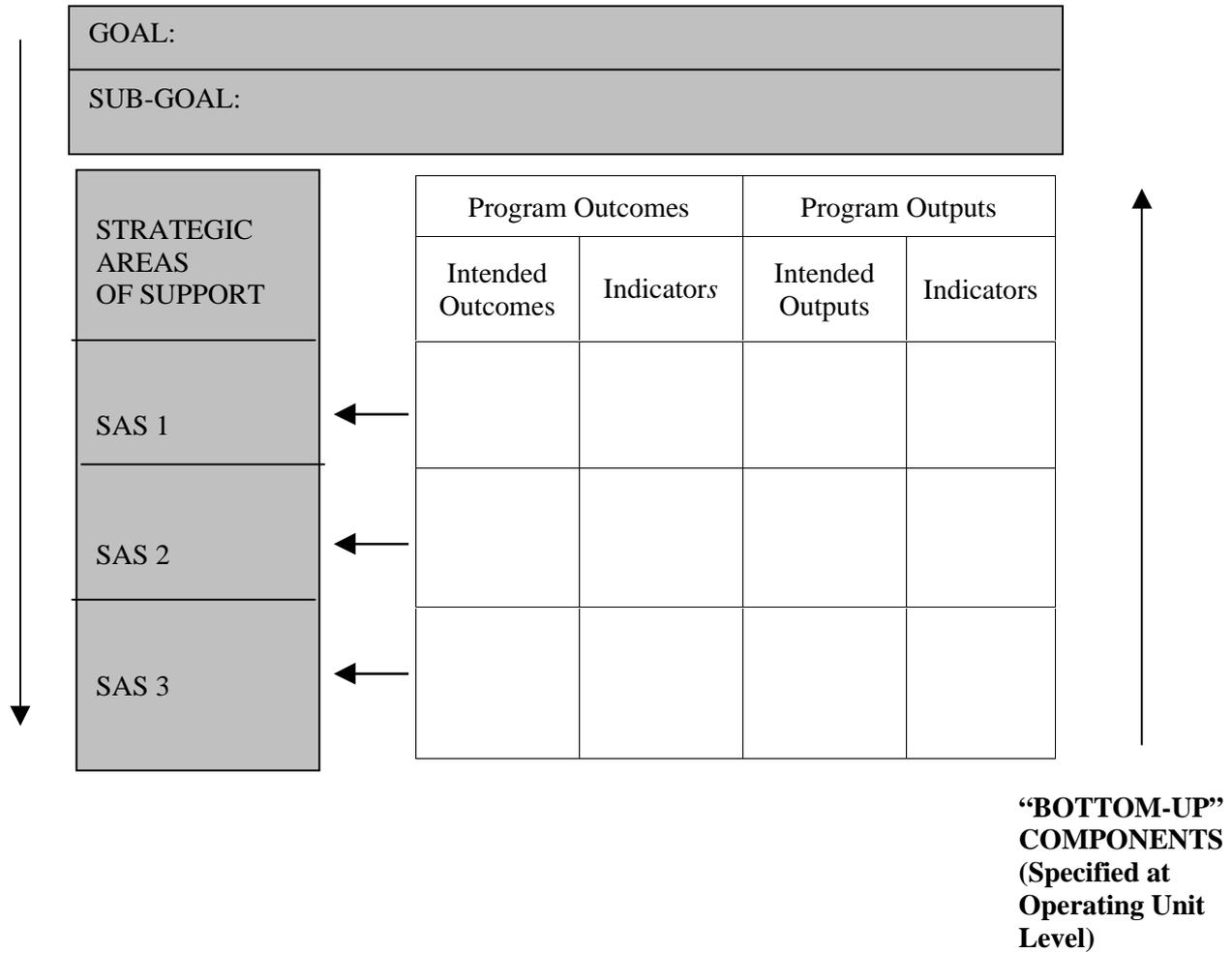
The UNDP allows its country operating units to select and report on their own (not standard) output and outcome indicators, but request the data be reported by strategic areas of support (equivalent to program approaches), which are corporately defined categories within the UNDP's strategic framework. Box 37 illustrates how the UNDP balances "top-down" versus "bottom-up" reporting relationships. Yet other agencies, such as AusAID, request that their country operating units report to headquarters on their own unit-defined project/program outputs and outcomes, categorized by agency-wide goals (called key results areas). These agencies allow their field units considerable flexibility in terms of what results they chose to report on, within headquarters-determined categories or structures. Once the data has been collected from the field, analysts in headquarters will – with the help of automated databases and sorting techniques – attempt to further categorize, analyze, aggregate and report on similar (common) types of outputs and outcomes within their agency's strategic framework structure. In other words, these agencies will decide *after* the output and outcome data are collected and entered into automated databases, exactly how it can best be sorted, categorized, analyzed, and reported.

It appears to be much easier to find standard or comparable indicators for project/program outputs than for outcomes, and output data is often more readily available from monitoring systems. Therefore, many donor agencies have started with efforts to aggregate output data. However, a potential shortcoming of reporting on project or program outputs (e.g., numbers and types of goods or services delivered, number of people reached) is that the response of intended stakeholder audiences may be "So what? What is the significance of this in terms of achieving significant development objectives?" Of course, as soon as one moves to higher-order outcomes or impacts, the issues of attribution and accountability begin to emerge. Some more immediate outcome measures, however, may be so closely or directly tied to a project or program that its contribution is obvious and doesn't require much further analysis. (For example, when trainees' test scores increase shortly after a training course). It is interesting that none of the donors have so far chosen to rely on standard customer surveys to collect comparable outcome-level information on customer satisfaction. Consistent data could be aggregated across projects; e.g., percentage of projects with customers who were fully satisfied (or better) with the project goods or services they received. Customer satisfaction may be viewed as a useful leading indicator, because unless they are satisfied, intended beneficiaries are unlikely to continue to use, and thus eventually be affected by a project's goods or services.

3.   Country-level development indicators. A number of donor agencies (e.g., USAID, DFID, World Bank) are using indicators and statistics on country-level development trends in the developing world in order to monitor results at the agency goal or sub-goal level. Such macro-statistics for developing countries are often readily available from international statistical sources (e.g., from UN organizations, the OECD/DAC, the World Bank and IMF). Mostly, these agencies have adopted indicators from among the core set of indicators agreed to by the international community (an effort co-ordinated by the DAC Working Party on Statistics) for monitoring country progress towards the shared goals and targets articulated in the report, *Shaping the 21st Century: The Contribution of Development Co-operation.* Examples of this type of indicator include infant mortality rates, school enrolment rates, percent of the population living in absolute poverty, etc. Not all donor agencies with strategic frameworks are using indicators of this type to report on their agency results, however. Some draw on these national, sector statistics more as "background" or "situational" indicators (e.g., UNDP, Danida), but keep them distinct from agency performance/results indicators, which are based only on their country offices' reporting of project/program outputs and outcomes.

Box 37: The UNDP's Approach to Country Operating Units' Reporting of Results within a Corporate Strategic Framework

"TOP DOWN" COMPONENTS
(Framework Specified by Headquarters)

GOAL:
SUB-GOAL:

STRATEGIC AREAS OF SUPPORT	Program Outcomes		Program Outputs	
	Intended Outcomes	Indicators	Intended Outputs	Indicators
SAS 1				
SAS 2				
SAS 3				

"BOTTOM-UP" COMPONENTS (Specified at Operating Unit Level)

Source: Adapted from *UNDP Results Framework: Technical Notes*, March 1999.

Reporting macro-statistics that show significant improvements in socio-economic conditions and welfare in the developing countries are likely to be important or valued in the eyes of domestic stakeholders and the public. The data are generally comparable (i.e., can be aggregated) across countries and are readily available via international or national statistical organizations. Using impact-level indicators drawing on national-level sector statistics *as measures of agency performance and results* has drawbacks, however. While more significant, the question that is then often raised is "Demonstrate how your agency's projects/programs have contributed to these overall results". Thus, attribution and related questions of accountability-for-results are usually the critical issues with this approach. Agencies can begin to narrow the range by limiting the country coverage to only those countries receiving substantial agency assistance (e.g., DFID's "top 30" approach), or better yet, by reporting only on those countries receiving the agency's project/program assistance in the relevant goal area (USAID's approach). Most agencies have attempted to get around the attribution and accountability issues by pointing out to stakeholders/oversight agencies that achievements at this level of impact are *shared or joint* responsibilities of the development community as a whole and cannot realistically be attributed to a single donor agency's assistance. Still, the questions of attribution and accountability may continue to be raised by auditors and other influential stakeholders, presenting a very challenging -- perhaps insurmountable – problem for those donor agencies who have chosen this approach. [31]

Examples of different donor approaches to selecting agency-wide performance indicators

The donor agencies reviewed have taken different approaches to the challenge of selecting indicators for measuring corporate-level performance and results. Below are some examples of their diverse approaches: [32]

- AusAID's approach will be to apply a set of performance indicators for reporting against each of its agency-wide goals (called Key Results Areas, or KRAs). The KRAs and their indicators are presented in AusAID's Performance Information Framework. Most of these indicators are generic or standard across KRAs – for example, expenditure ($m) by KRA, number of projects implemented by KRA, and percentage of projects rating satisfactory overall or higher within the KRA (based on AusAID's new project rating system). However, another planned set of indicators – significant project outputs achieved within each KRA – are not standardized or fully comparable across projects within KRAs. They will include number and types of outputs, and numbers of people assisted by these outputs, which will be sorted by headquarters into like categories/types of outputs (e.g., immunizations, training, micro-finance loans, etc.) after they've been reported by field offices. Initially, AusAID will not report on outcomes or impacts achieved at this time, although they will endeavour to expand to cover these aspects at a later stage. Annex 2.4 presents AusAID's Performance Information Framework and the indicators for each KRA.

- DFID has also developed a set of indicators to be applied to each of its four Agency goals (called objectives). DFID's approach involves an Output and Performance Analysis (OPA) which acts as a

---

31.   Previous chapters discussed how difficult it can be to convincingly demonstrate the net impacts or contributions of a particular agency project or program to national-level sector improvements in one country. More challenging yet would be seeking to demonstrate such contributions at a global/corporate level. Conducting a series of impact studies across all the countries receiving an agency's assistance within each goal area is a challenge of such magnitude and expense as to probably be infeasible. However, a few illustrative studies of this type might be enough to convince stakeholders that the agency's interventions are having a demonstrable impact.

32.   Agency approaches have been changing rapidly -- these are descriptions of approaches based on documentation available in the summer of 1999.

framework for measuring the agency's performance against its primary objectives. Only one of DFID's indicators in the framework is generic or standard across all the three goal areas. This indicator – the percent of projects likely to fully or largely meet their objectives – is based on DFID's project rating system. See Box 38 for details of how DFID is aggregating project ratings for agency-wide reporting and some of the issues encountered. The DFID's experiences with ratings are illustrative of similar approaches by other donor agencies. DFID's other performance indicators are goal-specific, country-level indicators related to "associated international development targets" (ITDs), as articulated in the DAC *Shaping the 21st Century* (and broadly agreed to by the development community and partners). In each objective area, DFID intends to track progress being made in the top 30 UK development partner countries, by using a few standard, national-level, sector-specific indicators that are for the most part related to the international development targets (IDTs). Annex 2.1 presents DFID's OPA framework, and indicators for each of the three objectives.

- Danida has taken a more detailed approach, preparing First Guidelines for an Output and Outcome Indicator System (September 1998) for country embassy reporting to headquarters on project outputs and outcomes according to specific indicators categorized by similar types of projects (i.e., program approaches). Within each of Danida's goal and sub-goal areas, the guidelines have identified standard indicators for each category of program approach. In other words, within a structured framework, standard indicators for similar types or categories of projects are used to collect, aggregate and report data on results (primarily outputs). Annex 2.6 illustrates Danida's output and outcome indicator system, by providing examples of indicators for selected program approaches. Danida has tested this new system in 20 countries. Data were collected and analysed and examples of the information were presented in Danida's first annual report:1998 (published in May 1999).

- USAID's indicator approach, as presented in its Agency Strategic Plan (September 1997), for reporting on Agency-wide performance and results within each of its seven goal areas, has been based on using country-level indicators of sector development trends. Many of these indicators are the same as those agreed to by the international development community in the DAC *Shaping the 21st Century*. Annex 2.2 presents USAID's strategic plan (strategic framework) goals, targets and associated indicators. USAID reported country trends data using these indicators in its last annual performance report and annual performance plan. Data was reported individually for each developing country receiving USAID assistance, and also aggregated in various ways (regional, global, stage of development, etc.). Recent experience with using and reporting data at this level is that it is technically easy enough, given readily available international databases and statistical software programs for data analysis. However, initial feedback (from USAID auditors and oversight agencies such as the OMB and GAO) has raised issues concerning USAID's accountability for these results and the agency's inability to demonstrate attribution of such high-order impacts to the USAID assistance program.33 USAID is now rethinking its approach for agency-level performance reporting, and may begin to draw more on its agency-wide program performance information database system (PMA) that gathers data on program results from its country operating units. Using a simple program rating system (based on a single criteria of effectiveness in achieving planned results), analysis has already been done of the percentage of USAID programs that have met or exceeded their targets, categorized within each of the Agency's key goal areas. So far, this data has not been widely disseminated for reporting performance at the agency-wide level, due to some

---

33. For example, a recent (April 1999) GAO review of USAID's *Annual Performance Plan for FY2000* listed the key weaknesses to be: "Does not develop clear linkages between agency and individual country goals" and "Does not identify the full range of other agency and other donor programs that may contribute to achieving the overall goals".

discomfort within the operational bureaus with the rating system approach. The rating system has been used internally, however, to aid in agency resource allocation decisions.

- The World Bank is developing a new Scorecard framework for measuring, assessing and reporting on corporate-level performance. The Bank's scorecard system is more complex than most other agency's strategic frameworks, involving several tiers or vertical levels at which performance will be measured and assessed. (See Annex 2.3 for a presentation of these tiers, their contents and indicators). Work is still on-going, in terms of selecting indicators and targets for some of the lower tier levels. The top level, Tier 1.A: Development Outcomes, identifies the Bank's four key development goal areas and the country trends indicators that will be used to track progress. The World Bank has selected eleven of the twenty-one core indicators from *Shaping the 21st Century* to monitor its development goals. While the scorecard is a new approach still under development, the Bank has for many years used project performance ratings (from implementation completion reports and periodic supervision reports) for annually aggregating and reporting performance across the Bank's project portfolio, much like the DFID and AusAID systems intend to do.

Box 38: DFID's Use of Project Ratings for Aggregating Performance to the Agency Level

DFID is in the midst of efforts to develop performance measurement systems that will enable performance reporting at the departmental and corporate levels. One aspect under development is the use of project ratings. Project Completion Reports are prepared for larger projects that assess what the project achieved and assigns scores. A recent synthesis of these reports found 64% of projects were considered likely to fully or largely achieve their objectives. Data on completed project performance ratings by DFID's new objectives framework is not yet available, because projects currently approaching completion pre-date adoption of the new objectives. DFID is now embarking on a system for scoring of larger projects under implementation via "output to purpose reviews" that require managers to annually assess progress towards delivery of outputs and achievement of purpose (outcome), and to assign performance ratings. DFID is also currently developing an automated management database system (PRISM) that will include a wide range of project information including performance ratings/scores. This system should facilitate aggregation across the project portfolio, providing a clear picture of departmental and agency-wide performance.

Operational departments have a number of concerns about scoring. There is a view that allocating a single score to complicated projects is questionable, and even more concern about aggregating scores from a portfolio of projects of varying complexity, which are often implemented in vastly differing country contexts. Importantly, if performance reporting is going to be used as a basis for future funding allocation decisions, there's a concern it may produce perverse incentives, such as encouraging project managers to opt for lower risk portfolios, even though riskier projects may have the greatest impacts on poverty. Such systems may be seen as putting too much emphasis on quantitative indicators and too little on qualitative factors. DFID is seeking to balance the needs of central departments for unambiguous performance data with the interests of the operational departments, by consulting them on system design to ensure it meets their operational needs.

Source: DFID, *Presentation by DFID at the Workshop on Performance Management and Evaluation*, October 1998; *An Approach to Portfolio Review in DFID*, 1999.

Phase 3: Setting targets

Setting targets (explicit indicator values to be achieved by a specific date) is particularly difficult at the level of agency-wide performance reporting. The related problems of attributing results and accountability-for-results, which are difficult enough at the project and country program levels, are greatly magnified at the agency-wide level.

Nevertheless, agencies are trying a few approaches to setting corporate targets. One common approach is to set targets (overall and within each key goal area) for the percent of project activities receiving satisfactory or better ratings based on project performance rating systems. For example, DFID has, for each of its goal areas, set a target of increasing the percent of projects likely to fully or largely meet their objectives to 75% by the year 2002. Similarly, AusAID has, for each of its key results areas, set a target of 75% of activities receiving a quality rating of satisfactory overall or higher for 1999/2000.

A number of agencies (e.g., DFID, USAID, the World Bank) have adopted the internationally agreed development targets articulated in *Shaping the 21st Century*, as their own (or with minor adjustments), integrating them into their agency strategic frameworks and plans. Usually caution is expressed that these are shared targets that are the joint responsibility of many development partners. The issue here is, will the oversight agencies and the public, to which the donor agencies are accountable, accept progress towards these shared targets as a reasonable approach to reporting on *agency* performance and results?

Monitoring the progress of developing countries towards long-term development targets (e.g., reducing percent of population in poverty by half by 2015, universal primary education by 2015, etc.) can be very useful as an agency strategic planning tool. For example, such data could be used for influencing agency resource allocation decisions on the basis of country needs (e.g., by ranking developing countries according to how far away they are from achieving the international goal targets). Moreover, monitoring these international development trends is important for assessing the effectiveness of the international development communities' combined efforts for achieving these shared targets. If sufficient progress is not being made, it should signal the donor community to re-think their development strategies. However, as an approach to targeting and reporting on a single *agency's* performance and results in the context of accountability, it may be open to criticism.

For example, in its first annual strategic plan, USAID mostly adopted targets that mirrored the internationally agreed targets, but opted in a few cases to set its own global targets (e.g., to reduce the fertility rate by 20% within a 10-year time-frame). In other cases, USAID established targets stated in annual terms (such as average annual growth rates in per capita income above 1 percent), or simply indicated the desired direction of change (such as primary school completion rates improved or loss of forest area slowed). [34] These approaches all suffer from the same problems of questionable attribution of national-level development trends to specific agency efforts, and how or if USAID can be held accountable for these results. USAID is now seeking other approaches, including more reliance on country program-specific results data reported by its country operating units.

Another approach, which was briefly considered by USAID and then (perhaps unfortunately) dropped, may still offer a different approach to setting targets at the agency level. Within key goal areas, specific *threshold values* for indicators could be established that would indicate when a country program is ready for graduation from assistance (i.e., further progress can be achieved by the partner country on a sustainable basis without outside assistance). This approach would have special appeal to legislators and

---

34.   This tactic has recently created some problems for USAID. Efforts to subsequently assess the Agency's performance in terms of these country- and global-level targets in annual performance plans has raised legitimate questions (by OMB, GAO, Auditors, etc.) about whether these targets are within the agency's sphere of influence and if it is possible to demonstrate attribution to USAID programs.

the public, since they hold the promise that programs will have an end in sight - once the development threshold values are achieved. For example, threshold levels for family planning programs might be when a country achieves contraceptive prevalence rates above 65% or when total fertility rates decline below 3.0. No specific deadline would be set for achievement of threshold values. This has the advantage of not holding an agency responsible for a specific rate of achievement at a level that is clearly well above its manageable interest (i.e., outside its sphere of control or influence), and yet may have appeal to stakeholder audiences. Agencies could report annually on the number of countries crossing key thresholds and thus graduating from sector assistance.

Phase 4: Collecting performance data agency-wide

Collecting data for agency-wide performance assessments and reporting takes place from two basic sources; (1) from existing international sources that maintain country level statistics on sector development trends, and (2) from the project/program performance measurement systems maintained by an agency's country operating units. Some agencies also synthesize key findings from a third source - evaluation reports. In most cases, these data are entered and stored in automated, central agency databases to facilitate agency-wide analysis and reporting.

Computer databases and software programs facilitate data sorting, aggregation, statistical analysis and graphic presentation of results. They can greatly aid the work of analyzing large amounts of performance and results data across project/program portfolios. Results of these agency-wide analyses of aggregate portfolio performance and results are usually reported to external stakeholders audiences in annual performance reports.

Central agency databases containing country development trends data

A few of the donor agencies reviewed have set up offices or contracts for collecting, storing and analyzing country-level data from international databases for annual agency performance reporting. For example:

- USAID has established a country trends database gathering national sector data for all USAID-assisted developing countries for each performance indicator in the USAID strategic plan/framework. This USAID automated database draws on other automated databases maintained by international statistical offices, non-governmental organizations, and the World Bank.

- The World Bank, in collaboration with the DAC Working Party on Statistics, has agreed to monitor developing country progress against the core set of indicators and goals/targets agreed to by the international community under the *Shaping the 21st Century* strategy. These data are published by the World Bank in their annual *World Indicators Reports* (also available on CD-ROM).

Central agency databases containing country operating units' performance information

Most of the donor agencies reviewed have also established or are in the process of establishing centralized, automated database systems for gathering, aggregating, analyzing and reporting data on project/program performance and results from their country operating units. Some examples:

- AusAID's activity management system (AMS) provides a centralized database with information for its project activities. AMS data includes financial and DAC sector coding information for each project activity, as well as project preparation and performance monitoring information (from activity preparation briefs and activity monitoring briefs). The AMS will incorporate the performance indictors for AusAID's new performance information framework, such as project

ratings and results (outputs initially, to be followed later with higher-order outcomes). Processes for data entry and for aggregating outputs from project to program and then to agency-level KRAs still need to be decided. The process for aggregating project ratings has already been established. Thus, the AMS will provide a standard reporting format for the monitoring and reporting of project activity performance and results. Moreover it will provide a mechanism for linking project activity and funding information to each of its KRAs (key goal areas), so that project expenditures, performance and results can all be reported by KRAs.

- DFID is developing a similar computerized project performance reporting system, known as PRISM, intended to facilitate the generation and analysis of information on the performance of DFID's project portfolio. PRISM, now in the pilot stage, will initially include primarily financial information and project performance ratings (based on annual scoring of on-going projects).

- USAID's program performance information system (called performance measurement and analysis or PMA Database) gathers country program results data (expected and actual results at the strategic objective and intermediate outcome levels) reported from its country operating units. PMA does not include information at the project level, nor does it incorporate financial/expenditure data.[35] PMA's results data are analyzed and used in USAID's annual performance reports and in technical annexes; e.g., to describe the agency's progress towards overall goals and to assess the extent to which operating unit programs are meeting their targets. Program performance ratings based on such analysis also have been used to inform agency resource allocation decisions across country programs.

Standard Data Entry Formats: To facilitate data entry, some agencies are developing standard formats for country operating units to report on their results. For example, UNDP recently has established a standard format for reporting country program outputs and outcomes within UNDP's new strategic results framework. Outputs and outcomes are reported at the level of strategic areas of support (i.e., program approaches). The UNDP standard data entry format is graphically illustrated in Box 37.

Danida also has developed a standard form for data entry into their results database. In Danida's system, country offices report results data (mostly outputs) for standard indicators by major program approach categories, identified in *First Guidelines for an Output and Outcome Indicator System.*

Marker Systems Linking Projects/Programs to Agency Strategic Framework Categories: An important aspect of these databases are the various "marker systems" or codes that serve to align or link individual projects or programs and their results according to the agency goals, sub-goals or program approaches to which they contribute. This serves to facilitate aggregation and analysis of project/program performance ratings and results data (outputs, outcomes) within the broader agency strategic framework structure. Exactly how and at which framework levels these "markers" (alignments) take place varies considerably from agency to agency.[36]

---

35. USAID maintains a separate central database for budgetary and financial information based on older, input-oriented activity codes. This older system has not yet been effectively integrated with USAID's program performance information database. This has hampered USAID's ability to relate agency results to their costs, and thus to establishing results-based budgeting approaches.

36. For example, in the UNDP and Danida systems, these linkages are made directly by country operating units when they enter project/program results data within standard reporting formats according to program approach categories specified by headquarters. AusAID's projects are linked via their DAC activity codes to the appropriate key goal areas (KRAs) in its AMS database. In USAID's PMA database, the operating units' strategic objectives are each assigned a code by headquarters which link them to the agency sub-goal (called agency objectives) to which they most contribute.

Moreover, some agencies (e.g., AusAID) are moving to integrate traditional activity codes and financial data with the performance data in order to facilitate accrual-based or performance-based budgeting.

Phase 5: Reviewing and reporting performance agency-wide

Options for aggregating performance and results agency-wide

Development agencies have a number of basic options to consider for aggregating or summing up performance and results achieved at the agency-wide or corporate level. At the two extremes – project/program outputs and country-level development changes (impacts) on the other -- aggregation of indicator data may be relatively easy. But in the case of outputs, the question "so what" may be raised. With country level development statistics, the key problem is that it is rarely possible to link changes credibly to a single agency's interventions, especially on a year-to-year basis. In the middle of these two extremes are project/program outcomes, which should be more significant than outputs (in the eyes of oversight agencies/ stakeholders), and yet have clearer linkages to agency activities than national development statistical trends. The problem here is that often there is such great diversity in projects' purposes and their outcome indicators that aggregating across projects or programs using comparable standard indicators is often not possible. Some agencies have overcome this problem by developing rating systems that score a project's success in meeting its objectives and then summing across projects the numbers and percentages that were successful or unsuccessful in achieving planned outcomes.

These options (and their pros and cons) are considered in detail below. They are summarized in Box 39.

1.   Selecting the project output level for agency-wide reporting on results. Outputs of projects (such as number of units of goods and services delivered or numbers of beneficiaries/clients reached) are generally easily collected and available annually. Moreover, outputs are relatively easy to sum up across similar types of projects. For development agencies with fairly centralized structures and a standard set of project outputs (goods/services) across country settings, this approach may be feasible. Danida is taking this approach, and has developed guidelines with standard output-level indicators for comparable types of projects within each of its major sectors. For agencies that have highly decentralized structures and great diversity in project types or approaches (e.g., USAID, UNDP), summing across project outputs may not be quite that easy. Another advantage of selecting the output level for agency-wide aggregation and reporting is that these results are easily attributable to an agency's projects/programs.

However, reporting at the output level will only be valuable to the extent that the intended external stakeholder audiences will be impressed with this level of results. If the response is "so what?", then summing and reporting on outputs may be of limited value in terms of defending the aid program.

2.   Selecting long-term country development trends data for agency-wide reporting of results. Another option for reporting on results achieved at the corporate level is to report on long-term social and economic changes at the country sector level using international statistical data, which have some measure of comparability across countries. Advantages of this approach include its appeal in terms of reporting on significant development impacts that matter to stakeholders, (e.g., alleviate poverty, reduce infant mortality, achieve universal primary education) and the ready availability of international indicator data covering many of the sector concerns of the development agencies. On the other hand, there are some serious issues with using this approach, especially in the context of accountability reporting on *agency* performance. Attempting to link and attribute these country-level and global-level development improvements to the activities of a single donor agency is often a wide stretch of the imagination that many will question.

USAID has attempted to use this approach recently in its initial response to GPRA reporting requirements in its annual performance plan and annual performance report. Despite attempting to place the country trends data in the context of joint responsibility and accountability with other development partners, USAID has still faced serious criticism from some stakeholders (e.g., IG auditors, GAO analysts). They want clearer accountability-for-results for which USAID is willing to be held responsible.

Another complication is that performance monitoring and reporting is usually done annually, whereas data on country development trends is often only available at intervals of several years and may not be very current. Moreover, even if data were available annually, the long-term nature of development impacts means year-to-year changes/improvements may not appear significant, and moreover cannot be attributed to current agency activities (due to means-ends lag effects).

3.   Selecting the project/program outcome level for agency-wide reporting of results. In between project outputs and macro-statistics, there's the level of project or program outcomes for which operating units are being held responsible. The outcome level is more significant (valued by stakeholders) than outputs, although perhaps not as much as developmental changes at the national level. Moreover, outcomes are more easily attributed to projects/programs that are country-level statistical trends, although not as easily as outputs. Another advantage is that performance monitoring systems at the project/program level are often already established, and thus some data on project/program outcome achievement should be available in most cases.

However, a key problem with aggregating project outcomes for agency-wide reporting is the typically great diversity of outcomes and their indicators, especially in decentralized agencies. Without standard indicators of project/program outcomes, direct aggregation is not possible. Some development agencies are now attempting to develop standard outcome indicators for common "program approaches" (groupings of similar projects). Other development agencies are getting around this incomparability problem by devising rating systems that score a project's success in meeting its outcome objectives. The agency can then aggregate across projects within a goal, sub-goal, or program approach area with statements like "85% of farmer credit projects successfully met or exceeded their outcome targets". Issues with this approach may include the extent to which standard criteria for making judgements about scores are applied across projects, and the reliability of "self-assessment" ratings especially when managers fear the consequences of poor scores. Moreover, it is uncertain the extent to which stakeholders will find reporting on project ratings satisfying.

Box 39: Pros and Cons of Different Options for Aggregating Results

Approach / option	Results Valued by stakeholders	Data Comparability	Data Available Annually	Results Easily Attributed
National Statistics (Development Impacts)	*high*	*high*	*low*	*low*
Project / Program Outcomes	*moderate*	*low*	*moderate*	*moderate*
Project / Program Outputs	*low*	*high*	*high*	*high*

Examples of donor agency approaches to assessing agency-wide performance

Several of the development agencies reviewed have had only limited experience with analyzing and reporting on agency-wide performance and results. Some have only recently completed development of their strategic frameworks that clarify their agency-wide goals, indicators and targets, and are still in the processes of developing databases and collecting data (e.g., DFID, AusAID, UNDP). In these cases, plans for exactly how data will be analyzed and presented in agency performance reports are still being developed. Others have completed one or more annual performance reports. Many donor approaches might be considered "experimental" and are still undergoing rapid change. Some examples of where donors are in the process:

- Danida has just recently completed its first round of data collection based on their new output and outcome indicator system. The data were analyzed and examples of the information were presented in Danida's first annual report, published in May 1999. As was expected, a number of methodological problems surfaced in particular sectors, and a major revision exercise in now going on.

- The World Bank is still developing indicators and targets for its new, comprehensive "scorecard" approach to corporate performance measurement, involving a complex framework with different vertical "tiers" (levels). Some implementation problems are apparent. However, the World Bank has a long history of aggregating, analyzing and reporting performance horizontally, across all projects completed in a particular year. For example, for more than twenty years, the OED has prepared Annual Reviews of Evaluation Results, based on aggregate analyses of project implementation completion reports and ratings. Recent reviews summarize the performance record for the projects completed during the previous year; examine the loner-term performance trends in terms of outcomes, sustainability, institutional development, and Bank performance; review performance vis-a-vis the Banks' key goal areas; analyze the influences on portfolio performance (including Bank and borrower performance); and explore policy implications and options for making further performance improvements.

- USAID also has considerable experience with preparing annual performance reports. The most recent reports and technical annexes draw on three central sources of performance information: (1) the performance measurement and analysis (PMA) database containing the operating units' program performance monitoring data (e.g., baselines, targets and actual results data for strategic objectives and intermediate results); (2) country development indicators (CDI) database containing country-level trend data for the indicators selected to monitor progress towards USAID's goals; and (3) the development experience system (DEXS) database which includes USAID evaluation reports prepared by the operating units and by the central evaluation office (CDIE). USAID's recent annual performance reports contain the following types of analyses: (1) a review of operating units' performance and results achieved during the last year; (2) a summary of findings and lessons from evaluation reports completed during the previous year; (3) an assessment of progress in implementing reform efforts under reengineering, especially towards results based management; and (4) a presentation of country trend indicators data for each of USAID's key goal areas. A supplementary volume, not widely circulated, analyzes program performance in key agency goal areas (based on a simple program performance rating system of whether targets have been exceeded, met, or un-met). In addition, USAID has in its annual performance plan attempted to report on annual performance towards the agency's key performance goals drawing on country development indicators, data and trends. This approach, as already noted, has been problematic and is being re-considered.

Performance budgeting

Performance budgeting (or performance based budgeting) is the general term used to refer to the infusion of performance information into resource allocation processes. Initially the effort is usually to link budgets with expected performance – that is, identifying what resources will be needed to achieve a planned result. (Eventually, the process may evolve towards allocation of resources based on actual, not just planned results. In other words, feedback on actual results achieved will influence future funding allocations towards better performing projects/programs).

Most OECD countries are now beginning to experiment with a variety of approaches to more closely link expected results with requested funding levels, as part of their broader reform efforts to become more results-oriented.[37] The concept of performance budgeting – essentially the process of linking budget levels to expected results, rather than to inputs or activities – is in most cases in early pilot phases and still evolving. No single definition or approach is likely to meet the varying needs of different agencies.

Most acknowledge that the budget process in government is essentially a political process in which performance information can be only one among a number of factors underlying decisions. Nevertheless, many governments have recognized that systematic presentation of performance information alongside of budgets can influence and improve budget decision-making.

A number of challenges face agency efforts to apply performance budgeting. Performance budgeting assumes that performance goals/ targets can be defined and that valid and reliable performance measures can be developed. Moreover, program outcomes and impacts are often not directly attributable to or under the complete control of single agencies. On the contrary, a donor agency's activities may be only a small component of overall public and private sector interventions in a given program area. In these cases, individual agency outcome or impact measures could be only of limited value to budgetary deliberations. Moreover, in development programs, desired impacts can only be achieved over many years. Relating these lengthy performance time horizons to annual budget deliberations can raise special measurement questions. In addition, the typical donor agency practice of using third parties (e.g., other government agencies, NGOs, contractors) as delivery or implementation agents mean these other partners' efforts, objectives and concerns are often critical factors in determining whether program results are being achieved, complicating agency accountability and attribution. Finally, the practice of earmarking funds limits the flexibility and ability of agencies to use performance criteria for resource allocation decisions and can cloud efforts to hold agencies accountable for results. For example, if funds are allocated to a country for political purposes even though it has an unfavourable policy environment that affects program performance, can the donor agency still be held accountable for the poor performance?

The high stakes involved in budgetary decisions further complicate the use of performance/results indicators. Introducing such measures into the resource allocation process heightens the potential for reporting biased or distorted data to make performance seem more favorable. Thus, procedures for verifying and validating performance data may become necessary.

---

37.   See OECD, *Budgeting for Results: Perspectives on Public Expenditure Management*, 1995; Julia Melkers and Katherine Willoughby, "The State of the States: Performance-Based Budgeting Requirements in 47 Out of 50", in *Public Administration Review* (Jan/Feb. 1998, Vol. 58, No.1); GAO, *Performance Budgeting: Initial Agency Experiences Provide a Foundation to Assess Future Directions*, GAO/T-AIMD/GGD-99-216, July 1999.

In addition to these problems with measuring performance/results, performance budgeting also requires an ability to accurately relate costs to results. Many agencies are now only beginning to develop adequate cost accounting systems that can be used to relate the full costs of various programs and activities to performance outputs.

Finally, performance budgeting almost always discloses tensions between budgeting structures and strategic planning frameworks. Agencies need to be able to link their new goal structures and their traditional activity structures (which form the basis of their budget requests). This is needed for straightforward connection among goals, budgets and performance information. However, achieving this link depends upon the capacity of agencies' program activity structures to meet dual needs. Program activity and budget structures generally evolved to monitor agency activities and spending and are geared to accountability for inputs and processes, not results. On the other hand, strategic frameworks need to be broad and wide-ranging if they are to articulate the broad goals that agencies seek to influence.

For example, agency goal frameworks that bear no connection to budget structures hamper performance budgeting initiatives. Agencies will need to bring these structures together, either by changing budget structures to more closely align them with agency goals/frameworks, or by using crosswalks or tables to show relationships between the two structures.

AusAID is an example of an agency that is undertaking a major integration of databases and re-coding exercise to enable the alignment or linkage of project activities, their results (outputs, initially) and their costs, to its new agency-wide goals (called key results areas). This will enable AusAID to report on numbers of projects implemented, total expenditures, and key results (outputs) against each of its key results areas.

USAID is an example of an agency that has not yet adequately connected its old project/activity coding system used for budgeting and financial accounting, with its newer program performance information system (PMA). While PMA's country program performance data from the operating units have been linked "upwards" (via markers/codes) to the agency's new strategic framework structure, it does not link "downwards" to individual project activities, their activity codes and funding data. This dis-connect hampers USAID's ability to report expenditures against agency goal areas, or to link costs to results/outcomes. USAID is working on this issue.

VI. DEFINING THE ROLE OF EVALUATION VIS-A-VIS PERFORMANCE MEASUREMENT

The role of evaluation vis-a-vis performance management and measurement systems is not entirely clear-cut, neither in the development co-operation agencies, nor more generally in the OECD country public sectors. Although evaluation was already a well-established practice when performance management systems were first introduced, in most cases the relationship between the two were not initially clarified. Perhaps this was because of some perceptions that evaluation had somehow failed to produce timely, relevant performance information and to influence management decisions, and that a new approach was needed. Also, evaluation may have been viewed as a specialized function carried out by external experts or independent units, whereas performance management and measurement essentially were seen as internal management functions.

Considerable effort is now going into clarifying the relationships, such as their distinctive features as well as their complementary nature. For example, at recent annual meetings of the American Evaluation Association numerous sessions were devoted to such topics. The U.S. GAO has recently prepared a brochure defining relationships between performance measurement and evaluation.[38] Several donor agencies have made efforts to clarify the respective roles of performance measurement and evaluation in various publications. (See Boxes 40-43.)

Before the rise of results based management, the traditional distinctions between monitoring and evaluation of development projects were fairly clear-cut. Monitoring was primarily concerned with implementation performance and compliance with planned schedules and budgets, whereas evaluation was focused on assessing higher-order development results. Moreover, monitoring was viewed primarily as an internal management function whereas evaluation was mostly external and independent to preserve objectivity. Monitoring took place routinely during implementation, whereas evaluation took place at specific times in the project life cycle – e.g., at mid-term, at completion, and occasionally ex post. These traditional distinctions may still be valid to some extent, but appear to be disappearing. For example:

- *The distinction that performance monitoring focuses on implementation while evaluation focuses on results is fading*. Under results based management, performance monitoring now emphasizes measuring results, including outcomes and even impacts. This new results-focus of performance monitoring systems is perhaps a response to the perceived failure of traditional evaluations to collect evidence on results. Similarly, evaluations also may focus on any of a broad array of performance issues, including implementation processes as well as results-focused issues such as effectiveness, impact, relevance, sustainability, etc.

- *The distinction that monitoring is an internal function whereas evaluation is an external, independent function also is diminishing to some extent*. Evaluation functions are increasingly being devolved to project/program management, although central evaluation offices often still conduct their own series of independent evaluations. Evaluations conducted or sponsored internally by management are believed to ensure greater ownership and use of evaluation findings for project/program improvements. Participatory techniques are also being encouraged to facilitate even greater involvement and ownership by various stakeholder and beneficiary groups. At the same time, there is growing concern over the validity, accuracy, and objectivity of performance monitoring data being reported by management self-assessments. More attention is being given to finding ways of validating, auditing, or independently reviewing performance monitoring reports.

---

38.   U.S. General Accounting Office, *Performance Measurement and Evaluation, Definitions and Relationships*, April 1998, GAO/GGD-98-26.

These trends highlight the tensions between the learning/improvement versus the accountability purposes of performance information, regardless of whether it is from performance measurement or evaluation sources.

- *Finally, both performance measurement and evaluation are increasingly being viewed as management decision-making tools whose timing should be geared to decision-needs.* Performance measurement data is gathered, analyzed, and reported routinely at specific intervals (usually annually) to provide "real time" feedback concerning performance. Whereas in some donor agencies evaluations are still planned or required at discrete times in the project cycle (e.g., at mid-term and completion), in other agencies, such as USAID, evaluations are now only conducted selectively when there is a specific management need. For example, when there is a need to explain unexpected results generated by performance measurement systems, or to analyze in-depth performance issues of particular interest to management. Rapid appraisal methods are also being encouraged in evaluations to ensure faster turn-around and thus relevance to management decision-needs. Evaluations as a matter of formality are being discouraged, whereas their potential as management tools are being encouraged.

- With traditional distinctions between monitoring and evaluation disappearing, how are these two functions now being differentiated under emerging results based management systems? Various donor agencies have defined performance measurement vis-a-vis evaluation functions in somewhat different ways. Different agencies tend to emphasize different distinctions. Some examples are given in Boxes 40-43. By no means are the distinctions yet clear-cut or broadly accepted. Generally speaking, though, there is an emerging agreement that they should be viewed as two separate approaches to producing different but complementary types of performance information – both of which are supportive of management decisions and thus equally important for effective results based management systems.

Both performance measurement and evaluation involve analysis and reporting on project/program performance and results. Evaluations are increasingly seen as more substantive, in-depth analytical efforts that can supplement the simpler forms of performance measurement analysis and reporting. Evaluations are undertaken when there is a need to better understand or explain project/program performance in its broader context or to generate recommendations for appropriate actions in light of that understanding. Moreover, whereas performance measurement analysis generally focuses on effectiveness (that is, whether results are being achieved as planned/targeted) and other simple performance measures, evaluation is better suited to address performance issues requiring more sophisticated methodologies (e.g., attribution).

Performance measurement and evaluation should also be viewed as inter-active and inter-dependent functions. For example, evaluators have often helped establish performance measurement systems; for example by training managers in self-assessment techniques and lending their methodological expertise to advise on various analytical matters (such as developing conceptual frameworks, selecting indicators and setting targets, collecting and analyzing performance data, etc.). Another inter-action between the two is that performance measurement systems serve as early warning to management when performance is unexpected, often triggering the need for evaluation to explain the gap in performance. Finally performance measurement systems can provide much of the results data that evaluators draw on and interpret, to arrive at judgements and recommendations for management action aimed at program improvement.

<div style="border: 1px solid black; padding: 20px;">

Box 40: CIDA's View of Performance Measurement and Evaluation

A recent guide on results based management in CIDA describes the differences between performance measurement and traditional evaluation as follows:

"Performance measurement differs from the traditional evaluation practice in that it is a continuous process of performance self-assessment undertaken by the program/project delivery partners. The traditional approach has been to schedule mid-term and end-of-term evaluations that are, generally, formative and summative in nature. These types of evaluations are typically conducted by external evaluators who are mandated to execute terms of reference set out by the funder which not only guide, but control the evaluation process. The evaluation exercise is often imposed on the other stakeholder groups as an administrative requirement. Because of the short timeframe within which to conduct these evaluations and lack of familiarity the evaluators have with the program/project implementation challenges, evaluations have tended to focus on management processes and not the achievement of development outcomes. Furthermore, evaluation recommendations are all too often written in an opaque manner so as not to offend the stakeholder groups. Evaluation research has shown that the utility value of traditional evaluations has been very low for program/project delivery partners and other stakeholder groups.

Within an RBM context, performance measurement is customised to respond to the performance information needs of program/project management and stakeholders. Since the stakeholders are involved in one aspect or another of measuring performance, the information that is generated is more accessible and transparent to the users. Performance measurement is also more results-oriented, because the focus is on measuring progress made toward the achievement of development results. Consequently, the performance information generated from performance measurement activities enhances learning and improves management decision-making."

Source: Quote from CIDA*, Results-Based Management in CIDA: An Introductory Guide to the Concepts and Principles,* January 1999.

</div>

Definitions

While there are no standard definitions distinguishing performance measurement from evaluation to which all donor agencies have agreed, some common themes have emerged which are summarized below.

Performance measurement involves routine or periodic monitoring, review (assessment), and reporting on project or program performance. It is particularly concerned with assessing effectiveness in achieving pre-established objectives and targets, although it also often addresses other performance issues as well. Assessments tend to be relatively straightforward presentations of results data or performance ratings following a standard format (often without a great deal of in-depth analysis, interpretation, context or explanation of the results presented). It is typically conducted as a self-assessment by project/program management, and is used both as a management tool and for accountability reporting to agency headquarters.

Evaluations can be defined as systematic analytical studies conducted occasionally or on an ad hoc basis, to answer specific management questions about performance. Evaluations may assess and explain any of a variety of project or program performance issues, but are particularly well-suited for dealing with more complex issues such as impact/attribution, sustainability, and relevance. They are often conducted by experts external to the project/program being evaluated, either from inside or outside the agency. But some may be self-evaluations conducted by project/program managers and may have participation by stakeholders and beneficiary groups. Evaluations tend to be more in-depth analyses that examine and explain performance in their broader contexts. They not only present evidence about results achieved (often obtained from performance measurement systems), but they interpret, explain, and make judgements about the performance in light of the conditions that influence the outcomes/impacts. Moreover, evaluations typically provide recommendations for actions to be taken that flow from their analysis. In other words, evaluations may draw their findings from performance monitoring results data, but go well beyond simple presentations of results, by drawing conclusions, interpretations or judgements based on an understanding of the broader context, and then making recommendations. Without an understanding of the underlying causes of performance shortfalls, which evaluations can provide, management may take inappropriate actions.[39] Moreover, evaluations often draw broader lessons for future project designs and/or for formulation of agency policies and program strategies.

Routine self-assessments by managers discussed in previous chapters (such as annual project performance reports, project completion reports, program results and resource request reports, etc.), as well as agency annual performance reports, are considered in this paper to be an integral part of the performance measurement system itself, and distinct from evaluation.

---

39.   Just knowing that a project/program has fallen short of its targets does not necessarily tell managers whether to terminate a program or increase efforts. For example, suppose an agricultural project aims to increase farmer incomes by increasing coffee yields, but is not achieving this goal. If the reason for a shortfall in performance is a drop in international export prices beyond the control of managers, it may be best to terminate the project, or switch to other less affected crops. If, on the other hand, the causes are lack of rural market access roads, an increase in resources directed at improving transportation may be the answer.

Box 41: USAID's View of Performance Monitoring and Evaluation

"*Performance monitoring systems* track and alert managers to whether actual results are being achieved as planned. They are built around a hierarchy of objectives logically linking USAID activities and resources to intermediate results and strategic objectives through cause-and-effect relationships. For each objective, one or more indicators are selected to measure performance against explicit targets (planned results to be achieved by specific dates). Performance monitoring is an on-going, routine effort requiring data gathering, analysis, and reporting on results at periodic intervals.

Evaluations are systematic analytical efforts that are planned and conducted in response to specific management questions about performance of USAID-funded development assistance programs or activities. Unlike performance monitoring, which is on-going, evaluations are occasional— conducted when needed. Evaluations often focus on why results are or are not being achieved. Or they may address issues such as relevance, effectiveness, efficiency, impact, or sustainability. Often evaluations provide management with lessons and recommendations for adjustments in program strategies or activities."

While performance monitoring and evaluation are distinct functions, they can be highly complementary if they are appropriately co-ordinated with each other.

Evaluations should be closely linked or integrated with performance monitoring systems. Performance monitoring information will often trigger or flag the need for an evaluation, especially when there are unexpected gaps between actual and planned results that need explanation. Depending on where the trouble lies, evaluations may be needed at the level of individual activities, intermediate results, or strategic objectives. Not only failures to achieve targets but also unexpected successes deserve special evaluations.

USAID operating units need to know not only what results were achieved (via the monitoring system) but also how and why they were achieved, and what actions to take to improve performance further (via evaluation). Thus, evaluation makes unique contributions to explaining performance and understanding what can be done to make further improvements. Evaluation is an important, complementary tool for improving program management."

Source: Quote from USAID, "The Role of Evaluation in USAID", *Performance Monitoring and Evaluation TIPS* (Number 11. 1997).

Evaluation issues

In performance management systems, the overall aim of evaluations is to answer specific management questions about project or program performance. The nature of these questions and thus an evaluation's focus may vary considerably from one to the next. Depending on their focus, evaluations may examine one or more of the following performance issues or questions.

- *Implementation performance.* Assessing specific implementation/process problems or the extent to which a project/ program is operating as intended – i.e., activities are implemented in conformance with program design schedules and budgets, statutory and regulatory requirements, etc

- *Effectiveness.* Assessing the extent to which a project/program achieves results (outputs, outcomes, or impacts) as planned or targeted. They usually also look at activity processes and context to understand how results were produced.

- *Unexpected results.* Answering questions about performance monitoring data – e.g., explaining gaps between actual and expected results by analyzing what factors may have impeded or contributed to its success.

- *Unintended results.* Analyzing a project/program's effects beyond its intended or planned results – i.e., unplanned results, positive or negative

- *Validating hypotheses.* Testing development hypotheses underlying logical frameworks – that is, assessing linkages between inputs, activities, outputs, outcomes and impacts and the intervening risk/contextual factors.

- *Customer satisfaction.* Assessing clients' or beneficiaries' perceptions about and responses to project/program services or products

- *Attribution.* Assessing the net outcomes or impacts attributable to a project or program. In other words, identifying the value added by a project/program to overall observed (gross) outcomes/impacts.

- *Efficiency, cost-effectiveness.* Comparing project/program outputs or outcomes to the costs (resources expended) to produce them, and identifying the least costly alternative to meeting a given result.

- *Cost-benefit.* Analysis of all relevant costs and benefits (usually expressed in dollar terms) of a project/program.

- *Relevance.* Reviewing the continued relevance of the project/program's results in light of changing beneficiary needs, partner country development priorities, or donor agency goals.

- *Sustainability.* Assessing the continuance of results after completion of a project/program (i.e., after donor support/funding terminates).

Box 42: The UNDP's Definitions of Monitoring and Evaluation

The UNDP's handbook for results-oriented monitoring and evaluation define monitoring and evaluation as follows:

"*Monitoring* is a continuous function that aims primarily to provide project management and the main stakeholders of an on-going programme or project with early indications of progress, or lack thereof, in the achievement of program or project objectives.

Monitoring enables a manager to identify and assess potential problems and success of a program or project. It provides the basis for corrective actions, both substantive and operational, to improve the program or project design, manner of implementation and quality of results…

Evaluation is a time-bound exercise that attempts to assess systematically and objectively the relevance, performance and success of on-going and completed programs and projects. Unlike monitoring, which must be undertaken for all programs and projects, evaluations are carried out more selectively for practical reasons. Programme or project managers have the flexibility to decide why and when an evaluation is needed based on a set of criteria.

If an evaluation is conducted at the mid-point of a program or project, it may serve as a means of validating or filling in the gaps in the initial assessment of relevance, effectiveness and efficiency obtained from monitoring. It may also assess early signs of project or program success or failure. If conducted after the termination of a program or project, an evaluation determines the extent to which it is successful in terms of impact, sustainability of results and contribution to capacity development."

Source: Quote from UNDP, *Part Two: Developing Selected Instruments for Monitoring and Evaluation.*

Box 43: The World Bank's Definitions of Monitoring and Evaluation

In a guide for designing project monitoring and evaluation, the OED provided the following definitions to distinguish between monitoring and evaluation:

"*Monitoring* is the continuous assessment of project implementation in relation to agreed schedules, and of the use of inputs, infrastructure, and services by project beneficiaries. Monitoring:

- Provides managers and other stakeholders with continuous feedback on implementation
- Identifies actual or potential successes and problems as early as possible to facilitate timely adjustments to project operation.

Evaluation is the periodic assessment of a project's relevance, performance, efficiency, and impact (both expected and unexpected) in relation to stated objectives:

- Project managers undertake interim evaluations during implementation as a first review of progress, a prognosis of a project's likely effects, and as a way to identify necessary adjustments in project design.
- Terminal evaluations, conducted at the end of a project, are required for project completion reports. They include an assessment of a project's effects and their potential sustainability.

Because *evaluation* is concerned mainly with impact, which will be measurable towards the end of implementation or in later years, it is often better done by a separate agency, independent from implementation. *Monitoring*, however, is a tool of good management, and the responsible unit should be located close to project management.

Source: Quote from the World Bank, "Designing Project Monitoring and Evaluation" in OED *Lessons & Practices,* June 1996, No. 8.

Distinctions and complementarities between performance measurement and evaluation

Performance measurement is commonly distinguished from evaluation in terms of different characteristics, focuses, and uses. However, due to variations among agencies, these distinctions may not be relevant to all agencies. Common distinctions between the two functions, as well as their complementary natures, are detailed below and are highlighted in Box 44:

- Performance measurement tracks whether or not results have been achieved as planned/targeted, while evaluations examine and explain why and how they were achieved or not. Thus, while performance measurement involves a straightforward assessment or observation of actual versus expected results, evaluation usually requires a more structured in-depth analysis of the context in which results were achieved.

- Performance measurement gathers and reports performance data routinely at specific intervals (e.g., quarterly, annually), whereas evaluation is conducted occasionally or selectively -- either at planned intervals (e.g., at project mid-term or completion), or on an ad hoc basis in response to management needs for specific performance information. (For example, to explain causes of an unexpected gap between actual and planned results/targets; to provide recommendations on how to improve performance; etc.)

- Performance measurement provides broad coverage by monitoring all key results, whereas evaluation typically provides in-depth analysis focusing on and explaining selected results of particular interest to management (e.g., those results for which performance is unexpectedly poor or successful). Performance monitoring reports are generally mandatory for all significant projects or programs, whereas evaluations are frequently optional and selective.

- Performance measurement and evaluation may both be asked to address any of the performance issues. However, performance measurement tends to focus on certain issues, whereas evaluation has advantages in other issue areas. For example, monitoring reports often focus on project/program effectiveness in achieving its development objectives (i.e., observing if targets are being met) or on simple measures of economy and efficiency. Evaluation is the better tool for assessing performance issues that involve more analytical effort and sophistication than routine performance monitoring and assessment can usually provide. For example, evaluations may focus on analyzing attribution, unintended effects, cost-effectiveness or cost-benefit analyses, relevance, sustainability, etc.

- Performance measurement usually involves quantitative indicators (even essentially qualitative information is transformed into numerical measures), whereas evaluation often relies on qualitative methods and analysis as well. For example, evaluation is increasingly emphasizing use of the more qualitative rapid appraisal techniques to save time and money. These rapid appraisals are especially appropriate for understanding the context in which performance takes place. Where possible, evaluations should draw on, analyze and interpret results data obtained from performance measurement systems. However, if evaluation is increasingly going to be expected to credibly assess performance issues such as attribution, supplemental fieldwork and perhaps more rigorous, formal methods may be required.

- Performance measurement is typically an internal, self-assessment tool employed by project/program managers, whereas evaluation is often conducted by an external, independent team (e.g., by a central evaluation office, external contractors, auditors or oversight agencies, etc.). However, evaluations may also be conducted internally by project/program managers, or in the case of participatory evaluations may even include other project/program stakeholders, partners and beneficiaries. Which evaluation approach is selected should depend on the purpose of the

evaluation. For example, if the key purpose of the evaluation is accountability, there will be greater need for external independence and objectivity of the evaluation team. On the other hand, if the primary purpose is learning and program improvement, a more internal or participatory approach is better.

- Performance measurement alerts managers to problems but usually does not tell them what to do, whereas evaluations typically make recommendations for action. In other words, performance measurement may be viewed as serving as an early warning system alerting project/program managers about unexpected results – but does not provide managers with appropriate solutions to the problem. In some cases, the causes of poor performance and the actions needed may be obvious, but in other cases they won't be and will require further in-depth analysis. Evaluations may be initiated to analyze the internal and external factors influencing and explaining performance and to make recommendations for actions flowing from their analysis.

Box 44: Key Distinctions between Performance Measurement and Evaluation

Performance Measurement	Evaluation
• Self-assessments by project/program managers	• Teams may have independent evaluators
• Broad coverage	• In-depth analysis
• Routine (e.g., annual)	• Occasional
• Mandatory	• Optional, selective
• Favors quantitative methods	• Often uses qualitative methods
• Answers what results achieved	• Answers why and how results achieved
• Alerts managers to problems	• Gives managers action recommendations

Potential conflicts and competition between performance measurement and evaluation

Despite the emerging view that performance measurement and evaluation are complementary and inter-dependent functions, and both equally essential to effective performance management, there is growing concern that they may be competing for the same, dwindling resources. The large efforts, in terms of staff time and other resources, needed to establish and maintain performance measurement systems and to report results may be drawing resources and staff away from evaluation. This may especially be the case in the context of overall reductions in government budgets and manpower levels. For example, U.S. federal evaluators surveyed in a recent GAO report on program evaluation raised this concern. (See Box 45.) Also, USAID has noted a sharp decline in the number of evaluations conducted by its country operating units since its "Reengineering" reforms in 1995 first required its country operating units to establish program performance measurement systems and made evaluations optional.[40]

---

40. The number of USAID evaluation reports submitted to the Development Experience Database dropped from 489 reports completed in FY1994 to 70 in FY1998. USAID has recently initiated a study of why the number of evaluations are declining and what to do about it.

Given limited and in some cases shrinking resources, agencies simply may not be able to establish and maintain new performance measurement systems and continue to conduct as many evaluations as before. To the extent that performance measurement also can provide managers with the appropriate information they need to make the right decisions, this might be ok. However, as already discussed, evaluation provides different, complementary kinds of performance information that may be equally important for making effective decisions. Thus, their potential decline should be viewed as a source of concern.

More effort within the donor agencies -- and perhaps among them -- needs to be given to positioning and integrating the performance measurement and evaluation functions vis-à-vis each other, and to enhancing their complementary natures in support of effective results based management.

Levels and types of evaluation

Below is a discussion of the key types of evaluations conducted by donor agencies at each of the three major organizational levels – project, country program, and agency-wide levels. Their complementary relationships (or potential relationships) with recently established performance measurement systems at each level is also examined. Many donor agencies are gradually shifting emphasis from traditional project level evaluations to broader country program and agency-wide evaluations. As this occurs, the value of conducting more joint evaluations with other donors and partners grows.

Project-level Evaluations. These evaluations focus on performance issues – either implementation/process issues or results-oriented concerns -- of a *single project*. Such evaluations may draw on project performance measurement systems for their data/evidence about performance achieved, but usually go beyond presentation of findings to explain the performance in terms of its broader context, and to provide recommendations and/or lessons learned. Project evaluations may address one or more of a variety of project performance issues, e.g. implementation processes and problems, efficiency, effectiveness, customer satisfaction, impact, sustainability, etc. Via in-depth analysis, project evaluations answer specific performance questions raised by management. They may, for example, investigate early warning alerts from project performance monitoring systems that performance/ results are falling short of expectations.

Box 45: GAO Surveys the Effects of Performance Management Reforms on Evaluation Function in U.S. Government Agencies

A recent report by the United States GAO surveyed evaluation units in U.S. federal departments and agencies to obtain their views about the likely impact that the Results Act -- introducing performance management as well as other on-going public sector reforms -- would have on the evaluation function. Generally it was felt that the Results Act, which recognized the complementary nature of evaluation and performance measurement, and specifically required summaries of evaluation findings in annual performance reports, would positively influence the demand for program evaluations. On the other hand, there was concern that the large effort required to establish performance measurement systems and produce valid and reliable results data would compete for funds currently used for more in-depth evaluations. Moreover, other reforms, such as reductions in staff and resources were likely to put added pressures on evaluation capacities and resources. Federal funding reductions were said by some evaluation officers to not only reduce evaluation activity but also reduce technical capacity, due to loss of experienced evaluation staff. Other reforms devolving federal program management responsibilities and accountability to states and non-governmental organizations were leading to increasingly diverse programs and complicating evaluation approaches and issues.

To meet anticipated increases in demand for program performance information as well as associated technical challenges, federal evaluation units were planning to leverage their resources by:

- Adapting existing information systems to yield data on program results.

- Broadening the range of their evaluation work to include less rigorous and expensive methods.

- Devolving program evaluation work to state and local program managers.

- Developing partnerships with others to integrate their varied performance information available on their programs.

- Increase technical assistance to state and local evaluators.

- Developing catalogues of tested performance measures for others to use.

- Conducting impact evaluations to supplement states' performance measurement information.

However, there was some concern that reliance on less rigorous methods would conflict with the need for accurate and credible information on program outcomes and impacts including attribution concerns.

The GAO study concluded that federal evaluation resources had an important role to play in responding to the increased demand for information on program results. They might best be leveraged by (1) assisting program managers to develop valid and reliable performance reporting under the Results Act; (2) planning evaluation studies to fill the most important performance information gaps, such as providing supplemental information on the reasons for observed performance; and (3) ensuring the results of diverse evaluation activities can be synthesized to portray programs at the national level by co-ordinating these activities in advance.

Source: U.S. General Accounting Office, *Programme Evaluation: Agencies Challenged by New Demand for Information on Programme Results*. GAO/GGD-98-53, April 1998.

Key types of project evaluations include:

- *Mid-term evaluations*: These are evaluations planned to occur during the implementation phase of a project. Their scopes of work may vary depending on the project's progress and management's key concerns. Generally speaking, they tend to focus on process, outputs, and short-to-medium term outcomes. Such evaluations draw on project performance monitoring system data on activity processes, outputs, and outcomes -- especially leading or proxy indicators, customer satisfaction surveys, etc. These evaluations are mostly conducted by project management or their contractors, and may include participation by other project stakeholders. Given that their primary purpose is usually project improvement, ownership by project management and key stakeholder groups is important. To some extent, mid-term evaluations may be being replaced in some agencies by new requirements that project managers prepare and submit annual "self-assessments" or project monitoring/appraisal reports (typically with performance ratings) to headquarters.

- *Final evaluations*: These are evaluations that take place around project completion, usually to assess performance in achieving project purpose (outcomes). They generally evaluate effectiveness in meeting outcome targets, but judge these achievements in view of the broader context of the project's environment as well as internal factors that may have influenced performance. Their purpose is not project implementation improvement (which is over), but may be for accountability purposes or to generate lessons learned for follow-on projects or for broader feedback and organizational learning. Final evaluations are separate in concept from project completion reports. Project completion reports are generally mandatory for larger projects (intended to provide complete project portfolio coverage) and are considered part of the performance measurement system itself. They are usually relatively straight forward, standardized report formats and ratings prepared by project managers. In contrast, final evaluations are usually conducted selectively, and are more in-depth studies often focused on specific performance issues. While final evaluations may draw on performance monitoring system data, they generally involve additional field-work and analysis by an evaluation team (which may include project managers but also may include external evaluators, contractors, and/or other stakeholders).

- *Impact and other ex post evaluations*: These evaluations are timed to take place several years after project completion. They are usually conducted selectively, as part of series, by central evaluation offices. They are generally far less in number than mid-term or final evaluations. Ex post evaluations tend to focus on achievement of long-term development results or impacts, and are often concerned with attribution (net impact) or sustainability issues. Given their timing and focus, they are not initiated or conducted by project managers but by higher organizational levels, such as by the country operating unit, or by the agency's central evaluation office (often as part of larger sector program impact studies). Data from broader program performance measurement systems that extend beyond the individual project's life cycle may be helpful to such studies. Alternatively, impact evaluations may have to mount their own data collection efforts.

- *Ad hoc/special evaluations*: It may be worth noting another category of ad hoc or special evaluations. Whereas mid-term and final evaluations are typically planned in advance (and may be requirements in some agencies), special evaluations are more ad hoc in nature and are typically initiated by project managers to address specific performance issues or problems. For example, they may be conducted in response to early warnings from project performance monitoring systems alerting managers that targets are not being met. USAID no longer requires conducting evaluations as a matter of formality, instead favoring ad hoc evaluations timed to assist managers with specific information needs for decision-making.

Country-level program evaluations: Country program evaluations (also referred to as country sector assessments, sector impact evaluations, or strategic evaluations) are focused on the combined performance of a *group of related projects or other activities* aiming at the same development objective within a developing country. In other words, they assess the effectiveness of a set of related activities in terms of achieving specific country-level development results, usually of a sector or sub-sector nature. Such evaluations may attempt to compare and assess the relative effectiveness of the different project intervention strategies aimed at the same objective, their synergies and potential conflicts or tradeoffs. Such evaluations would find program performance monitoring data based on relevant country results frameworks especially useful to analyze, if they exist. There are several basic types of country program evaluation:

- *Single donor country program evaluations (sector-specific)*: Such evaluations focus on the effectiveness of a single donor agency's program assistance strategies (i.e., related sets of agency projects and activities) that contribute to a single sector or sub-sector development objective in a country. An example would be the World Bank OED's new pilot series of broadly conceived sector impact evaluations that assess the impact of all Bank assistance in a given sector and country, including non-lending work, during a certain period of time. These evaluations are also intended to examine recent projects or those still undergoing implementation, and to assess the relevance of the current strategy in light of past experience. Very similar in concept are USAID's traditional program impact evaluations conducted by the central evaluation office (CDIE). Country operating units that have adopted the country program approach to performance measurement would benefit from conducting occasional strategic evaluations focused on the validity of the development hypotheses underlying their results frameworks, and the effectiveness of their intervention strategies.

- *Single donor country assistance evaluations (multi-sector)*: Related in concept, but broader, are evaluations that assess the effectiveness of a single donor agency's overall, multi-sector assistance efforts within a country. An example is the World Bank's recently initiated series of Country Assistance Reviews (CARs). CARs focus on the impacts and effectiveness of the Bank's country assistance strategies towards achieving its priority development objectives -- i.e., across several key sectors, rather than just one. While the World Bank seeks team participation by other donors in conducting their CARs and usually includes a section discussing donor co-ordination, their focus is mainly on the Bank's program strategy and performance.

- *Joint program evaluations (sector-specific)*: These evaluations are equally concerned with the effectiveness of different donor agencies' and partner organizations' activities and programs all aimed at achieving a shared development objective within a country. The evaluations examine and compare the effectiveness of the different partners' programs and strategies, their co-ordination, and possible duplications of effort. These evaluations are jointly conducted by the different donor agencies and partner organizations involved. An example is the joint evaluation of the Rwanda emergency relief assistance program conducted a few years ago. Members of the DAC Working Group on Aid Evaluation have recently discussed the advantages of conducting a series of such joint program evaluations. The first has recently been completed for the transportation sector in Ghana. Such evaluations might draw on and benefit from performance monitoring data based on joint country results frameworks, where they exist.

Agency-level evaluations: These evaluations generally focus on a donor agency's world-wide or sometimes region-wide performance in:

- Key sectors or sub-sectors (e.g., primary health care, small-scale enterprise, farmer credit).

- Cross-cutting themes (e.g., participation, gender).

- Policies or operational systems (e.g., design approaches, partnership practices, budgeting systems).

Usually they are conducted by a central evaluation office or perhaps by an evaluation unit attached to a regional, or technical/sector department. Several methodologies and data sources may be used for these agency-level evaluations. They may, for example, review and synthesize findings from a series of existing evaluation reports conducted by operating units on related projects/programs in a given sector or theme area. Centralized databases with performance measurement system information would also be likely sources of information for agency-wide evaluations. For example, project completion reports may be used as building-blocks for agency reviews/evaluations. So might the newer, annual project/program performance reporting systems and databases now being developed by some agencies. International statistical databases with country trends indicators might also be drawn upon. In addition to reviews of existing databases and reports, another typical method involves conducting a series of more in-depth, field-based country case studies that are representative of typical agency program strategies and performance. From these case studies an overall synthesis report is then prepared.

Multi-donor, world-wide evaluations: A final category of evaluations are those conducted jointly by a group of donor agencies that synthesize performance and lessons learned world-wide, usually focused on a specific sector, sub-sector, cross-cutting theme, or operational system of particular interest to the donor group. These evaluation studies generally review and synthesize experiences across the donor agencies. Examples include the evaluation syntheses conducted by the DAC Working Party on Aid Evaluation on topics such as participatory development and good governance, women-in-development, sustainability of primary health care programs, results based management, and many others.

Conclusions

To some extent, the evaluation function and practices of the donor agencies have continued along traditional lines since the introduction of performance measurement systems, although in some cases with less emphasis, priority and resources than before. New performance measurement and reporting tasks may compete for the same resources and staff effort formerly devoted to evaluation. Some efforts have been made in most agencies reviewed to distinguish the roles of evaluation vis-a-vis performance measurement, to better define their complementarities, and to understand the unique contributions each can make to results based management.

Nevertheless, confusion still exists within many donor agencies, especially at operational or procedural levels. More work remains to be done in terms of re-positioning evaluation so that it is better integrated with - and thus can contribute more to - newly established performance management systems. This may involve revising agency evaluation policy and practices guidelines to better tailor it towards the newly established results based management systems, policies and procedures. Also, the central evaluation office may need to better plan its own evaluation agenda as well as co-ordinate the evaluation plans of operating units so that evaluations will contribute more directly to results based management decision-making processes at different organizational levels and to external reporting requirements (e.g., annual performance reports).

Steps might be taken to better integrate these centralized evaluation efforts with newly established performance measurement and management systems. For example, central evaluation offices might:

- Focus their central evaluation agenda topics on key agency goals or sub-goals identified in their newly established agency strategic frameworks.

- Synthesize findings from operating units' project/program evaluation reports in order to help meet annual agency performance reporting needs.

- Co-ordinate the evaluation plans of the country operating units so they might better address new agency goal areas and performance reporting needs.

- Share their evaluation staffs' methodological and conceptual expertise by assisting in agency efforts to construct strategic frameworks, develop performance indicator selection criteria, recommended standard indicators, or other aspects of establishing performance measurement systems.

- Provide training in basic evaluation concepts and techniques to project/program managers who are increasingly responsible for analyzing and reporting on their own performance.

- Assist in efforts to review or validate management self-assessments.

- Effectively implementing results based management at higher organizational levels implies a shift in attention from traditional project level evaluations to more emphasis on country program evaluations and agency-wide evaluations. As this shift takes place, the benefits of conducting joint evaluations with other donors and partners will grow substantially.

Finally, donor agencies could share and learn more from each other's experiences in terms of how to effectively position the evaluation function within emerging results based management systems. Moreover, there may be advantages to seeking greater harmonization among donor agencies' in their performance measurement and evaluation terminology, definitions and distinctions, in the context of results based management. Further co-ordination of PM&E approaches and requirements would not only benefit the donors, but perhaps even more so, the partner developing countries who must deal with the confusion and overlaps at the country level. However, the extent to which donor agencies can co-ordinate may be limited by their need to follow government-wide mandated approaches.

VII. ENHANCING THE USE OF PERFORMANCE INFORMATION IN THE DEVELOPMENT CO-OPERATION AGENCIES

This chapter reviews the key uses of performance information in the donor agencies. In results based management systems, performance information is not collected for its own sake, but for continuous feedback into management learning and decision-making processes that will further improve projects and programs and accelerate achievement of results. In addition to this internal management learning/improvement function, performance information is also used to fulfil several accountability functions. Within the agency, project/program managers and operating units are increasingly being held accountable for achieving and for reporting results to higher organizational levels. Moreover, donor agencies are increasingly being required to report results to external domestic stakeholder audiences (legislatures, oversight agencies, auditors, the public, etc.). Donor agencies also have a special obligation to report on performance to their partners – developing countries and beneficiary groups. Sometimes these various uses appear to be in conflict with each other, raising the question of how or whether a single performance information system can respond adequately to all needs. This chapter also briefly explores various steps being taken in agencies to enhance the use of performance information.

Since most donor agencies are in the early stages of establishing their results based management systems, especially at the higher organizational levels, their actual experience with performance information use and with mechanisms for enhancing its use are still limited or not well documented. Thus, this chapter outlines intended uses, initial practices and experiences, and early impressions.

Key uses of performance information

In results based management systems, performance information (from both performance measurement and evaluation sources) serves two primary aims or uses. One use is as an internal management tool for making project and program improvements; the second is for accountability reporting.

- **Management improvement (managing-for-results)**: The first major use of performance information is to provide continuous feedback to managers about the results they are achieving, so they can then use the information to improve their performance even more. This use is often referred to as "managing-for-results". Sometimes discussions of this internal management use are further sub-divided into two related aspects or processes -- promoting learning and facilitating decision-making.

    1. *Promote learning.* Performance information promotes continuous management learning about what results are being achieved by their projects/programs and why --i.e., what factors are influencing good or poor performance. Improved knowledge is a prerequisite for better decisions.

    2. *Facilitate decision-making.* Management's learning in turn facilitates their making appropriate decisions. Continuous performance information about progress towards results and about the factors influencing performance will facilitate good decision-making and timely action. Lessons from experience can help agency managers to continually improve their development assistance

projects and programs, and to formulate better policies. A special type of decision-making that performance information is increasingly being called upon to influence is resource allocations.

- **Performance reporting (accountability-for-results)**: The second key use of performance information is for performance reporting and accountability. Accountability-for-results has several dimensions. One distinction is between external (organizational) accountability and (internal) individual or work unit accountability. In the case of donor agencies, external accountability involves not only domestic stakeholders, but extends to development partners. Performance reporting/accountability use can be sub-divided into the following categories:

- *Accountability to domestic stakeholders*. Donor agencies, like other domestic government agencies, are accountable for achieving and reporting results to their taxpaying public and elected representatives, and to designated central oversight/auditing agencies. The donor agencies reviewed are committed to publishing annual performance reports for these external domestic audiences, transparently reporting on the performance and results achieved by their development assistance programs. Often there are now government-wide legal requirements or executive orders for reporting results, at certain times and in specific formats, which are being audited by oversight agencies.

- *Accountability to development partners*. In addition to being responsible to their domestic stakeholders, the donor agencies are also accountable to their various development partners – e.g., developing country governments, implementing partners, other donors, and ultimately their intended beneficiary groups.

- *Internal management accountability*. Moreover, agency accountability for achieving and reporting results is increasingly being devolved and translated into accountability at lower organizational levels (e.g., operating units, teams, or even individual managers). Several agencies are experimenting with management contracts and personnel appraisal systems that specify what results are to be achieved, when, and by whom.

Potential tensions between uses

As experience with using performance information grows, the potential for tensions or tradeoffs between its two primary intended uses (management improvement versus accountability reporting) is emerging. From a management improvement perspective, one challenge of results based management has been to shift focus from inputs to outputs and from outputs to even higher outcomes and impacts. Not only is it important to know what results are being achieved at these different levels, but also to understand the cause-effect linkages between them -- e.g., why an activity is successful or not, which approaches work better, and under what conditions or contexts. An audit/accountability mentality, however, may ironically shift focus back down to outputs, which can be more easily attained and attributed to agency activities, and for which data can be easily collected on an annual basis. Managers have relatively greater control over outputs and thus are understandably more comfortable with being held accountable for this lower level of results than for outcomes or impacts over which they have less and less influence. Moreover, outcomes and especially impacts are longer-term changes that may not show improvements quickly or annually. Since performance reporting is conducted annually, this is encouraging managers to search for lower-level results to report, which would show changes faster. Moreover, there is a growing concern among auditors/oversight agencies with attributing results to particular agency interventions. Since demonstrating attribution becomes increasingly difficult for higher-order outcomes and impacts, this also acts to encourage managers to focus and report at lower results levels. Furthermore, accountability reporting tends to emphasize measuring what is being achieved (and comparing it to pre-set targets), rather than analyzing why or how it is being achieved. In contrast, a management improvement/learning approach is equally

concerned with analyzing the context and factors influencing performance, and with drawing lessons for improving performance.

Accountability reporting versus management improvement uses also implies different data collection and analysis approaches. For example, concerns over attributing outcomes and especially impacts to specific agency interventions imply highly rigorous designs and data collection methods. It also implies extensive attention to data quality, validity and reliability, and to independent verification. On the other hand, a management improvement approach would tend to emphasize more rapid and low-cost data collection/appraisal techniques, with data of sufficient quality for decision-making needs but not necessarily up to standards required for social science research. Moreover, it would favor a self-assessment approach to ensure management's ownership and a first-hand learning experience, and also would encourage more participatory methods and stakeholder involvement. In contrast, an audit/accountability approach might either call for more independent assessments or for a system of spot-checks, reviews and verification of management self-assessments.

Finally, these two intended uses may influence agency management's behavior differently. While a managing-for-results focus would tend to encourage risk-taking, experimentation and learning, there is a danger that an emphasis on accountability-for-results might encourage more conservative or risk-averse behavior (e.g., avoiding potentially risky projects, focusing on lower-level results, setting easily attainable targets).

These conflicting aims present a dilemma for donor agencies, as it does for other government agencies implementing results based management. Both primary uses need to be kept in mind when establishing performance measurement and evaluation systems. To the extent possible, the systems will need to address both uses and mix or balance data collection and analysis approaches to satisfy both interests. For example, an independent series of central impact evaluations might be undertaken to address auditors' concerns about accountability and attribution, while managers in the field might be encouraged to conduct self-assessments employing more rapid appraisal and participatory techniques. Box 46 highlight some key tensions between accountability-for-results and managing-for-results uses of performance information.

Box 46: Common Tensions between the Key Uses of Performance Information

Accountability for Results	Managing for Results
• Emphasizes meeting targets	• Emphasizes continuous improvements
• Focus pushed down to outputs	• Focus shifts up to outcomes and impacts
• Requires independent assessment or verification	• Emphasizes self-assessments and participation
• Greater concern with attribution	• Less concern with attribution
• Implies rigorous methods and high quality data	• Favors rapid, low cost methods
• Encourages conservative behavior	• Encourages risk-taking, experimenting, learning

It may be worth mentioning at this point a few refinements. For example, the use of performance information in resource allocation is included under the internal management improvement category because it involves a decision-making process. However, the resource allocation use is really rather different from the management learning/improvement function that takes place at the project/program level. These functions typically involve different users, for different purposes, with different motivations and perspectives. The former is more concerned with what happened in terms of results, whereas the latter is more concerned with why. The former employs a more ex post, summative approach, while the latter puts more emphasis on learning while a project/program is ongoing. The resource allocation use, while internal to the agency, is external to line management, and an element of defensiveness or even hostility can creep into the dialogue.

Another point worth making here is that a tension may exist between uses of performance information for reporting to domestic stakeholders versus the interests of partner developing countries. To the extent that the various donor agencies must follow government-wide performance reporting requirements, this may limit their ability to harmonize their performance measurement and reporting needs with those of their partners in developing countries. Developing countries may be faced with potentially competing and overlapping frameworks, methodologies, data and reporting formats of different donors. Rather than assisting and supporting developing countries' capacities to generate and use performance information, donors may actually be over-burdening those capacities to meet their own domestic reporting needs.

More detail on the uses of performance information at various agency organizational levels follow. Also, care should be taken to guard against the possibility of misusing performance measurement information. Box 47 addresses a few of these inappropriate uses.

Box 47: A Few Examples of Inappropriate Uses of Performance Measurement: Guarding Against Distortion

Performance monitoring data can usefully serve to motivate project staff to achieve results. This influence, embodied in the saying "What gets measured gets done", can be very positive. However, there is a possible danger that unless the indicators selected are valid and capture all key aspects or dimensions of the result being sought, staff seeking to improve performance against specific measures may actually distort a project's true objectives. An example would be when family planning workers provide contraceptives to women who have passed their child-bearing age in order to inflate their performance measures. Another example would be where targets for achieving a high numbers of outputs might cause staff to ignore quality concerns. This problem is not insurmountable, however, and can be addressed with use of multiple indicators and qualitative studies as well as adequate supervision.

Another potential problem is embodied in the adage "You become what you measure". The UNDP, for example, is particularly concerned that their areas of comparative advantage, the so-called "soft" areas of capacity-building, policy advice and advocacy are the most difficult against which to assess results. The experience of a number of donor agencies with the shift to a results based management approach has shown that, unless guarded against, there could be a tendency to focus on activities/approaches that are easily quantified. The World Bank's recent policy research report, *Assessing Aid: What Works, What Doesn't, and Why* makes precisely the point that assistance relating to the policy environment and institutional capacities can have a far greater impact than aid to specific measurable sectors with clearly defined goals.

Performance information uses at the project level

Performance information at the project level is used for two main purposes:

- Learning and making improvements to on-going projects.
- Accountability and reporting on project performance to higher organizational levels.

Learning and making improvements at the project level: At the project level, performance monitoring information is used to provide continuous feedback to project staff and managers concerning their implementation progress and results achieved compared to targets. This can act to motive them and also serve as an early warning when plans are not being achieved. Supplementary evaluation studies can help managers to better understand the causes of problem performance, where it is not obvious from monitoring data alone, and recommend what corrective actions may be required to put the project back on track.

The types of performance information needed will change as the project matures. In early phases, information on inputs and implementation processes will predominate. This will give way to a focus on outputs and shorter-term outcomes during the middle years, and eventually shift to medium-term outcomes at the project's end, and ultimately to long-term impacts ex post.

Moreover, performance information needs will vary according to different management levels. For example, project field staff will be most concerned with performance data about implementation processes, middle management may concentrate on achieving output targets, and higher project management levels will also want to know about outcomes and achievement of project purpose. Impacts, which are typically only observed after project completion, is primarily of interest to agency senior managers at higher organizational levels. Generally, people are most interested in, and have the most use for, information concerning their own performance – in other words, at the level for which they are responsible and accountable, and over which they can exercise some control and decision-making power.

Accountability reporting on project performance to higher organizational levels: In addition to using the performance information to learn and to take corrective actions, project staff and managers use their performance information systems to report their performance to higher organizational levels and to various stakeholder groups, thus serving an accountability function as well. For example, implementing agency field staff will report on their progress to their project supervisors/managers; they in turn report the project's performance to the sponsoring donor agency operating unit, and perhaps also to other partners and stakeholders, e.g. government officials and beneficiary groups.

Performance information uses at the country program level

Main uses of performance information at the country program level include:

- Promoting learning and influencing country-level programming and resource allocation decisions.
- Reporting the operating unit's project/program performance to agency headquarters.
- Sharing performance information and collaborating with other partners.

Promoting learning and influencing country programming and resource allocation decisions in the country operating units: At the level of country operating units, performance information is used to make broader programming decisions. For example, knowledge gained about what works and what does not from previous projects may feed into the design of new follow-on projects. In some donor agency operating units, annual or bi-annual reviews are held in which performance data across a whole strategic

development objective area are discussed and assessed. In these review sessions, the comparative performance of different activities aiming at the same objective are assessed. This may lead to shifting resources from the less successful to the more successful activities. It is useful to bring the cost factor into the equation, although not all agencies do this. In other words, it is not just which activities are producing more results that matters, but which are producing them more cost-effectively. The criteria for comparison becomes which activity approach is producing a given result at least cost.

When reviewing performance information and making decisions at the country program level, it is also important that operating units analyze contextual data or draw on supplementary evaluations that help provide a better understanding of the factors influencing program performance and achievement of objectives. Applying too rigid a formula of shifting resources from poorer to better performing activities may not be advisable. In some instances the solution for poor performance may be more funding, not less, whereas in other cases where the cause of poor performance is beyond agency management's control, then terminating the project or reducing funding may be appropriate. Other factors besides performance, such as the priority placed on aiding a specific target group in need, or the political importance of a project or sector, must also be weighed in programming decisions. Bringing other development partners and stakeholders into the operating unit's review process will help broaden the discussion, increase co-ordination efforts, and make the programming decisions more participatory.

Reporting the country operating unit's project/program performance to agency headquarters: Country operating units also use program performance monitoring data and complementary strategic program evaluations to report on and explain their program performance to agency headquarters. For example, USAID operating units report annually to headquarters on their program's actual results vis-à-vis their strategic objective targets in the R4 reports. These reports also estimate the unit's resource requirements needed to achieve future planned results. The R4s are reviewed jointly between the country unit and the regional bureau at USAID headquarters once per year. At the reviews, joint decisions are made concerning program performance and priorities, resource allocations, revisions to strategic plans, data collection and quality issues, etc. Headquarters also uses the operating units' performance information in R4 reports for compiling into agency-wide performance reports.

Sharing the operating unit's project/program performance information with partners and harmonizing approaches at the country-level: Country operating units may also use performance information to report on their activities' performance to the partner country government, other donors and stakeholders, to which they are also, in a sense, accountable. The concept of joint accountability for achieving a shared development objective in a country implies at a minimum a sharing of performance information among partners involved. More ambitiously, it implies a participatory process in which partners might use performance information about their respective interventions in order to take co-ordinated or joint programming and resource allocation decisions aimed at improving performance of a country program for which the partners are jointly accountable.

Conflicts among uses and users may arise in this regard. To the extent that different donor agencies may already have developed different country program approaches (e.g., different frameworks, terminology, indicators, analytical methods and reporting formats), the partner country government will to deal with trying to co-ordinate and make sense of these different and possibly competing approaches. As already discussed under previous chapters, greater donor co-ordination of performance measurement and reporting systems at the country program level would be very advantageous for all agencies trying to work together in-country, and especially for the partner country government. However, in-country harmonization efforts may be constrained to the extent that donor agencies must follow the dictates of their domestic government-wide performance measurement and reporting requirements.

Performance information uses at the agency level

Key performance information uses at the agency level include:

- Preparing agency annual performance reports for external stakeholder audiences.

- Influencing agency resource allocations across countries and sector programs.

- Contributing to organizational learning and improving agency policies and program strategies.

Preparing Agency Performance Reports: The primary use for performance information at the agency-wide level thus far has been for annual performance reporting to various domestic stakeholder audiences. The diverse analytical approaches to external reporting taken by the donor agencies has already been reviewed in the previous chapter.

Agencies may have a number of aims for their external reporting; for example, to "sell" their development assistance programs to the public and to their elected representatives; to convince these audiences that they are achieving worthwhile results with the taxpayers' money, and to successfully compete for additional funding. Other domestic stakeholders that must be satisfied include those oversight agencies tasked with auditing compliance with government-wide legal requirements for reporting results (such as the General Accounting Office (GAO) in the U.S. or the Office of the Auditor General (OAG) in Canada). Most donor agencies have sought to meet these various stakeholder interests and demands with the same annual performance report. However, to the extent that reporting to meet legal requirements or to satisfy particular auditor issues may not have broad appeal, perhaps separate reports directed at different audiences or purposes might make more sense.

Influencing Agency Resource Allocations: There is not a great deal of information on how the donor agencies intend to use or are using performance information for agency-wide resource allocation progresses. But some experience is available. For example, in USAID, program performance information reported by the operating units has been used to influence resource allocations across country programs, with the general aim of shifting resources to better performing programs that are meeting or exceeding their targets and away from those that fail to meet their targets. A review of this resource allocation process last year concluded that other factors besides the program performance ratings tended to predominate, and performance criteria appeared to make a difference only in marginal cases. Whether USAID will continue to rate and rank-order program performance as a factor in budget allocation decisions or revise the procedure in future years is now being reconsidered. The process is more fully discussed in Box 48.

The concept of factoring program performance into resource allocation decisions is fundamental to results based management. Using performance information to influence resource allocations among activities aimed at the same development objective *within* a country is a useful practice. However, donor agencies need to consider several issues or problems with allocating resources *across* countries and programs based on performance information. First, how practical or just is it to compare performance or results achieved across widely different country contexts? Program performance is greatly influenced by a country's level of development – poorer countries tend to be poorer performers because of lack of basic institutions, capacities, and their general poverty situation. In other words, allocating funds across countries on the basis of program performance may be at odds with poverty or need criteria; it may end up diverting assistance away from the neediest countries.

Box 48: USAID's Use of Performance Information in Budget Allocation

In last year's annual budget allocation processes, USAID's regional bureaus reviewed all of their operating units' R4 reports, and assessed the performance of each strategic objective (program) according to whether it met, exceeded or fell short of its expected targets. This performance score comprised 50% of a final composite ranking for each strategic objective. Other factors besides performance that also influenced the composite ranking included country development considerations (e.g., country need and quality of the development partnership); the contributions of the strategic objective towards U.S. national interests in the partner country, and the relative importance of the country to U.S. foreign policy concerns). Regional bureaus then ranked similar strategic objectives within each Agency goal area into three groups: the top 25%, the middle 50% and the bottom 25%. This approach identified those objectives (programs) with the most compelling case for funding, those with moderate justification, and those with the least. The rankings were used to influence resource allocation decisions among country operating units and their strategic objectives (programs) within the regional bureaus. A recent review of the process concluded:

> "USAID is utilizing performance to inform budget allocations. By and large, good performance is being rewarded; the better performing programs are receiving greater budget allocations and poorer performing programs are receiving less monies. In addition the regional bureaus are using technical reviews to evaluate missions' self-assessments of performance, and using common factors and weights ... to rank strategic objectives for budget allocation purposes. The presence of earmarks and directives, and the need to be responsive to concerns outside the Agency, however, limit the Agency's ability to use performance to inform budget allocations".

Approaches to further streamline this process and make it more uniform across bureaus is being considered. However, others have recommended dropping the process altogether, as being too time consuming given its marginal influence in a system predominated by earmarks and political considerations.

Source: USAID/PPC, *Performance and Budget Allocation* (September 1998, draft).

Moreover, applying performance criteria according to a rigid formula (e.g., good performance automatically equates with more money, whereas worse performance means less) may not be appropriate. The *causes* of poor program performance may differ widely and needs to be considered in any funding decision. In some cases, the appropriate solution for lagging program performance may be to increase funding, not decrease it. Finally, if it becomes widely known that resource allocations are being rigidly or mechanistically tied to performance, this may lead to incentives for system distortions. Program managers and operating units will find it in their interest to lower their targets, distort performance data, and put their performance in the best possible light in order to avoid budget cuts.

One alternative possibility donors may wish to consider is using country-level statistics for the internationally agreed development indicators and targets expressed in the DAC *Shaping the 21st Century* report as criteria for allocating their development assistance resources across countries. The DAC core set of indicators could have great value for agency-wide strategic planning purposes – such as for developing objective criteria to influence resource allocations among countries based on their comparative need, or for deciding which countries might "graduate" from program assistance based on their meeting or surpassing performance targets or thresholds. For example, for a particular DAC goal/indicator, developing countries

could be rank-ordered according to their current status – i.e., whether they fall below, meet or already exceed an international target or threshold value. Those countries already exceeding the internationally agreed target might be considered candidates for graduating from assistance in that sector/goal area. Countries with large gaps or falling way below the target/threshold levels would receive greater assistance given their greater needs.

Contributing to Organizational Learning and Improving Agency Policies and Strategies: A final use of performance information is to promote agency-wide organizational learning about what works and what does not, under what country contexts. Often this is a key function of central evaluation offices. Donor agencies may learn important lessons in this regard by analyzing project performance databases and synthesizing project completion reports or annual progress reports, for example, as the World Bank does in its *annual review of evaluation results* and *annual report on portfolio performance*. Performance monitoring reports at the project level may be used as building blocks for analyzing various performance criteria with regard to the agency's overall portfolio, or within particular sectors, sub-sectors or thematic areas. Similarly, project evaluations may be synthesised, although it should be remembered that their coverage is usually less complete and their formats less consistent. Another approach frequently taken by central evaluation offices are to conduct a series of field-based project or program evaluations within a particular sector, goal or thematic area, and then prepare a synthesis drawing on them as case studies. All of these approaches have common aims -- not only of assessing performance in a particular area, but also of understanding the factors and contexts that influence that performance, in order to draw lessons with relevance for agency-wide policy and program strategy formulation. Knowledge about what works and what does not under what conditions can then be translated into agency policies and program strategy planning and "best practices" guidance for managers.

Organizational Approaches for Encouraging the Use of Performance Information

Some agencies have also planned or instituted other significant organizational reforms in support of results-oriented management. Many of these changes act to stimulate or facilitate the use of performance information. Some of the key reforms include:

- *Decentralization.* This involves delegation of authority over project activities and resources to the country operating units and down to the management level that can best achieve results in the field. "Letting the managers manage" involves giving lower management levels the flexibility they need for promptly shifting from poorer to better performing activities. It may also involve other reforms, such as organizing "strategic objective teams" empowered with authority and resources to achieve specific development results, as opposed to more traditional organizational structures and management hierarchies. It also implies shifting greater responsibility and authority to implementing agency partners.

- *Accountability*. Along with delegation of authority empowering managers to achieve results comes greater accountability. New mechanisms are being instituted in several donor agencies for holding their senior and even line managers accountable-for-results, such as results-oriented management contracts and personnel appraisals. Care must be taken so that these mechanisms provide positive incentives for managing-for-results, and not disincentives that may lead to undesirable risk-averse behaviors or to dishonest or distorted reporting of results. Some agencies are exploring the concept of joint accountability with other partners for achieving higher-order results. Also, autonomy and accountability must go hand-in-hand. Insufficient autonomy to make decisions will eventually undermine accountability and create frustration among managers and staff where new requirements are imposed without sufficient resources and flexibility to meet them.

- ***Client focus.*** Most donor agencies are increasingly consulting with beneficiary groups in partner developing countries concerning their preferences and satisfaction with project goods and services provided, and are being more responsive to their needs and desires. Some have instituted polices and procedures requiring country operating units to develop customer service plans and conduct periodic customer surveys, consultations or assessments. Satisfying customers (intended beneficiaries) in developing countries is one aspect of a donor agency's accountabilities under results based management. Moreover, unless customers are sufficiently satisfied, it is unlikely they will continue to use project/program outputs (goods, services), and thus higher-order outcomes and impacts will not be achieved.

- ***Participation.*** The inclusion of development partners and stakeholders in all aspects of the performance measurement and management processes is policy in all the donor agencies reviewed. All donor agencies reviewed have policies supporting participatory processes for jointly setting objectives, defining indicators, collecting data, analyzing and reviewing data, conducting evaluations, and using the information in decision-making. Participation in these processes improves partner and stakeholder ownership of project/program objectives, indicators and targets, and builds support for performance measurement systems and their use. A number of constraints, however, have limited actual practices.

- ***Reformed policies and procedures***. Several of the agencies reviewed have developed and published new policies and procedural guidelines for changing the way the agency conducts its business. These directives generally spell out how the new results-orientation will be operationalized in agency processes and defines new organizational roles and responsibilities for its implementation. For example, guidance addresses new requirements for strategic planning, performance measurement and evaluation systems, and outlines new institutional procedures for routinely using performance information in agency decision-making and budgetary processes, and in external reporting requirements.

- ***Supporting Mechanisms***. In addition to declaring new policies and procedures, most agencies have sought ways of supporting their managers' efforts to follow the new guidance and establish effective performance measurement and management systems. These diverse efforts include: (a) major "re-engineering" training exercises; (b) establishment of central units responsible for providing technical assistance to decentralized units; (c) preparing supplemental technical references or "best practices" guidance materials on various aspects of establishing results based management systems; (d) building central databases with consistent, well-organized, timely, and easily accessible performance information; (e) developing conceptual frameworks, planning and reporting formats, and various other tools, tips and techniques designed to assist managers to effectively implement performance measurement and management systems. Many of these mechanisms have already been discussed in this paper or illustrated in the various boxes.

- ***Cultural change.*** There are two key aspects to managing organizational change. The first deals with the external changes in the way an organization does business, such as establishing new systems, structures, roles and responsibilities. Equally important for successful implementation of results based management is making changes in organizational culture - that is, transitions in the internal values, feelings, attitudes and behaviors of an agency's personnel. For example, results based management requires a new emphasis on learning and integrating lessons into decisions, a focus on achieving results rather than implementing inputs and processes, and a new commitment to open and transparent performance reporting. It also implies paying attention to helping people cope with change, by dealing with their internal feelings and emotions. Cultural changes, such as those implied by new results-oriented management systems, might be supported by staff training, knowledge networks, help desks, and counselling services, in order for the new ways of doing business to be effectively and enthusiastically institutionalized.

VIII. CONCLUSIONS, LESSONS AND NEXT STEPS

This final chapter offers some concluding remarks about the state-of-the-art of results based management and remaining issues or challenges facing donor agencies. Also provided are some preliminary lessons being learned regarding effective practices for establishing results based management systems. These comments are based on the review of donor agency documents, the discussions held at the DAC Working Party on Aid Evaluation's Workshop on Performance Management and Evaluation (New York, October 1998), as well as a brief review of the broader literature on performance measurement and management.

Plans for a second phase of work on results based management (agreed to by the Working Party on Aid Evaluation at their meeting on 10-11 February 2000) are also briefly outlined.

Some conclusions and remaining challenges

The OECD countries are increasingly facing "aid fatigue" and there are growing pressures on donor agencies to show development results. This is part of a much broader trend within the OECD countries to reform their public sectors to make them more effective and performance-oriented. Stakeholders want and expect the donor agencies, like other domestic government agencies, to be accountable for and report on results accomplished with taxpayers' money. In response, many donor agencies have been establishing performance measurement and management systems to complement their more traditional monitoring and evaluation systems.

Donor agencies face special challenges in developing effective performance measurement and management systems that are different from, and in some ways more pronounced than, the challenges faced by most other domestic agencies. For example, donor agencies must work in many country settings and across many sectors. Their products and services are often more diverse and finding comparable indicators that can be aggregated across programs and countries is difficult. Moreover, donor agencies typically are not just doing simple service delivery, where results are relatively easy to measure, but instead do a lot of institutional capacity-building and policy reform, which are less easily measured. The donor agencies' performance measurement systems are also constrained by their dependence on the typically weak institutional capacities of their partner countries for collecting results data.

Progress and experience with results based management systems differs considerably from agency to agency. Some donor agencies reviewed (e.g., USAID) have accumulated nearly a decade of experience with implementing performance measurement and management systems. However, most are still in early stages of developing their systems.

Generally speaking, most donor agencies have a tradition of monitoring performance and using such information at the project level. However, most emphasis has previously been on monitoring project *implementation* performance -- that is, with tracking inputs, activities and processes. With the introduction of results based management, however, there has been a shift in emphasis to monitoring project *results* – that is, outputs, outcomes and impacts. There are differences among donors in terms of the level of results emphasized. Some donors (e.g., USAID) have focused mostly on monitoring higher-order outcomes and impacts, while others (e.g., Danida, AusAID) have at least initially focused their systems on tracking immediate project outputs. A challenge for donors will be finding the time and resources to "do it all" -- that is, balance project performance monitoring needs at all logframe hierarchy levels, without overburdening the monitoring system or having it displace evaluation or implementation activities.

Some donor agencies are also beginning to develop performance measurement and management systems at higher organizational levels – above the traditional project level. USAID has perhaps the most experience with establishing systems at the country program level, although there are significant efforts going on in other donor agencies as well. For example, the World Bank's Comprehensive Development Framework, piloted in twelve developing countries, provides a useful organizing framework for how development partners ought to conduct development efforts at the country level. Working at the country program level implies following principles of partner country leadership, participation and partnership among development partners, and a shared development strategy and results-orientation. The country program approach to performance measurement and management is particularly well suited to newly emerging modes of assistance such as joint sector investment programs. While most donor agencies support these principles as policy, there are numerous constraints that still limit the extent of their actual practice.

All the donor agencies reviewed are now feeling pressures to report annually on agency-wide results to their external domestic stakeholders and taxpaying public. In some cases, such reporting is now a government-wide requirement under law or executive order. Thus, performance measurement and reporting at the overall agency or "corporate" level is becoming an urgent priority. Efforts to measure and report on performance at these higher organizational levels raise the twin challenges of *aggregating* and *attributing* results. *Aggregating* results refers to how the donor agencies can best "add up" or summarize their results from lower organizational levels – i.e., from projects or country programs -- to agency-wide, global levels. *Attributing* results refers to convincingly demonstrating that they are the consequence of an agency's interventions and not of other extraneous factors. Attributing results convincingly becomes more difficult as one moves beyond immediate outputs to intermediate outcomes to long-term impacts. Moreover, it becomes progressively more difficult as one moves from the project level, to the country program level, and ultimately to the global level.

Domestic stakeholders may be asking for inherently conflicting things in agency annual performance reports. Not only do they usually want to hear about results that are developmentally significant, but also about results that can show annual improvements and can be attributed to an agency's own projects/programs. The former implies monitoring higher-order outcomes and impacts while the latter implies tracking more immediate outputs. Moreover, the results data needs to be relatively comparable across projects or country programs, which would argue in favor of choosing the extremes of either impact or output level data, and against selecting intermediate outcome data, which are generally more diverse. Some donors have also tried to enhance comparability of results either by developing menu systems of "common indicators" (i.e., comparable or standard indicators) or by establishing performance rating systems. None of these approaches appears clearly superior to others at this point, and for now donors may be best advised to continue experimenting and mixing these approaches. More work needs to be done to develop methodologies for aggregation and attribution of results that will be convincing and credible to stakeholders but also will be within reasonable levels of cost and effort.

A related challenge is seeking to strike a balance between top-down direction from agency headquarters and bottom-up flexibility for field managers in terms of results reported. Reporting results at the corporate level requires a clarification of overall agency goals and the development of a framework that facilitates measuring and aggregating results globally within these goal areas. Therefore, some direction and structure from headquarters is necessary. On the other hand, there are dangers in designing performance measurement systems too much from the top-down. Unless there is a sense of ownership or "buy-in" by project/program management and partners, the performance data are unlikely to be used in operational decision-making. Moreover, imposed, top-down systems may lack relevance to actual project/program results, may not sufficiently capture their diversity, and may even lead to program distortions as managers try to do what is measurable rather than what is best. Field managers need some autonomy if they are going to manage-for-results. Some operational level flexibility is needed for defining, measuring, reporting, and using results data that are appropriate to the specific project/program and to its country setting. Striking a balance between a headquarters determined structure (needed for aggregating and

reporting results) and field unit flexibility (for determining what results are appropriate in a given context) is another key challenge facing donor agencies. Different donor agencies have been approaching this issue differently, with different degrees of centralization.

The donor agencies (and also OECD public sector agencies more generally) have most experience with implementing performance measurement systems, some experience with reporting performance, and least experience with using performance information for management improvement. In other words, documented examples of the actual uses of performance information in management decision-making processes in the donor agencies are still limited. Also, not much has been written about experiences with specific mechanisms, procedures and incentives these agencies may have adopted to stimulate demand and use of performance information for management learning and decision-making purposes. More attention to these areas and identification of "best practices" would add significantly to the current state-of-the-art.

There is growing evidence that the two primary intended uses of performance management systems – that is, (a) for external accountability reporting - holding an agency and its managers accountable for results, and (b) for influencing internal management learning and decisions -- may not be entirely compatible with one another. To some extent, they imply focusing on different types of results data and alternative methodologies, and may even influence management incentives and behaviors differently. There is growing concern whether and how results based management systems can simultaneously and adequately serve *both* of these key uses -- particularly when the external reporting function appears to have the greater urgency. Avoiding overwhelming the system with external demands, and keeping an appropriate focus on internal *managing-for-results* uses is yet another challenge facing the donor agencies.

Also worth noting, donor agencies are different from most public agencies in an important way. In addition to the usual accountability to their *domestic* stakeholders and public, donor agencies also have unique responsibilities to *foreign* stakeholders and publics – that is, to the partner developing country agencies and beneficiary groups with whom they work. The types of performance information most suitable for the domestic versus partner country audiences and users may be different, placing additional, conflicting demands on performance measurement and management reporting systems. For example, from a partner country's perspective, the ideal would be for all donors/partners to support a jointly-owned performance measurement system, tailored to the specific country program and context. However, for donor agencies to aggregate and report results across countries and programs, performance information must be comparable rather than country-specific. Moreover, government-wide reporting requirements influence a number of donor agencies' reporting systems, which further constrain their flexibility to co-ordinate at the partner country level.

One of the key decision-making processes that performance information is intended to influence is agency resource allocation. Initially, performance budgeting involves estimating the budget requirements needed to achieve specific planned results (project/program outputs, outcomes, impacts). However, this may not be as simple as it seems. Because traditionally budgets were linked to inputs or activities, linking them now to results may require changes in financial accounting practices and coding systems. A number of the donor agencies are now making such accounting changes to enable better linking of resources (expenditures) with planned results. Moreover, if the results are *outcomes or impacts,* there is the additional issue of the extent to which these results are attributable to specific project/program expenditures.

At later stages, performance-based budgeting may also involve the re-allocation of resources according to *actual* results achieved. In other words, budget decisions may be influenced by actual performance data, so that resources are shifted towards better-performing activities and away from poorer-performing activities. The influence of such performance criteria on the budget allocation process across countries is likely to be limited, given the largely political nature of budget decisions, the existence of other legitimate criteria (e.g., country need, capacity, and commitment), legislative earmarks, etc. Generally speaking (with the exception of USAID), there is not much practical experience yet available in the donor agencies with using

performance information as criteria in the across-country resource allocation decision-making process. It is likely that performance budgeting may be more appropriate for allocating resources among projects/programs *within* a partner country, rather than *across* countries.

Donor agencies have made some progress in clarifying the respective roles of performance measurement and evaluation. Most agencies view them as distinct functions offering complementary types of performance information – both of which are considered important for effective results based management. Whereas performance measurement answers what results were achieved and whether targets were met, evaluation explains why and how these results occurred, focuses on understanding successes and failures, and draws lessons and recommendations. Evaluations may also look beyond effectiveness (whether intended results were achieved), at issues of relevance, attribution, cost-effectiveness, sustainability, and unintended results. Performance monitoring is generally viewed as the responsibility of project management, while evaluation may be conducted by external teams, by internal management, or by mixed teams. The evaluation function may be the responsibility of an independent evaluation office or may be internal to project management, or a mix of both. External evaluations may also serve to validate the findings of performance monitoring reports, which are self-assessments.

However, the specific distinctions made between performance measurement and evaluation functions vary somewhat from agency to agency. For example, the donor agencies vary in terms of the degree of importance placed on "independence" of the evaluation function.

While in theory both performance measurement and evaluation functions are now seen as critical to effective results based management, in practice there are signs that these functions may actually be competing for the same scarce resources and staff time. For example, in USAID the number of evaluations have dropped significantly since the mid-1990s when performance measurement and reporting procedures were first required agency-wide. Recent U.S. GAO reports indicate there is a growing concern among U.S. government agency evaluators about how the Results Act is affecting them, as evaluation budgets and staffing levels continue to decline. Donor agencies may be challenged to protect their evaluation activities from being overwhelmed by new performance measurement and reporting requirements.

A final set of opportunities and challenges facing the donor agencies are concerned with co-ordination of performance measurement and management efforts. Harmonization among the donor agencies, at least initially, need not necessarily mean adopting a standardized approach or sharing common performance information systems and databases. Rather, they might start with donors adopting similar terminology and definitions for the same concepts, and by sharing information concerning effective practices, useful tools and methods. There would certainly be advantages to sharing a common vocabulary, building on each other's experiences, and avoiding duplication of efforts. The DAC Working Party on Aid Evaluation's October 1998 Workshop on Performance Management and Evaluation and this state-of-the-art review of donor experiences are first steps towards such sharing of experiences. The DAC Working Party on Aid Evaluation has also made progress on a *Glossary* exercise that compares various donors' terminology and definitions for both evaluation and results based management concepts. The glossary demonstrates the great diversity that currently exists. A second phase of work on results based management that will identify and share "good practices" among donor agencies has been initiated by the DAC Secretariat. Even such relatively simple steps towards co-ordination are not without challenges, however. For example, once agencies have already "invested" in their own particular terminology and definitions, it may be difficult for them to agree to change to a common set of definitions and terms. In addition their freedom to harmonize terminology or even to adopt the "best practices" of others may to some extent be constrained by government-wide directives that dictate their approach.

Considerable progress towards performance measurement harmonization has also taken place among donors on another front. Broad, sector development goals, targets and indicators for measuring progress at the country and global levels have already been adopted by the donor community in the shared strategy,

Shaping the 21ˢᵗ Century: The Role of Development Co-operation. The DAC Working Party on Statistics has been at the forefront of this effort to identify and seek agreement on a core set of indicators for measuring performance vis-à-vis the shared goals. Several of the donor agencies reviewed have linked their own corporate strategic framework goals and indicators to these internationally agreed goals and indicators. Now, a more concerted effort by the donor community is needed to support partner countries' capacity to collect data and monitor progress towards the international goals over the coming years.

An even more ambitious form of harmonization among donors and partners might be envisioned for the future in the context of country development efforts. Donor agencies' performance measurement and management systems need to be co-ordinated at the country program level in a way that lessens the burden on the partner countries' capacities, builds their ownership and leadership capabilities, and enhances the development of sustainable results based management systems within partner countries.

Such a co-ordinated country-level approach would reduce the potential burden on partner countries' capacities of having to deal with diverse and even competing performance information needs and systems of the different donor agencies. Sector program performance monitoring and evaluation systems could ideally be owned and maintained by the partner country to meet its own information needs on development results, with the donors being supportive, assisting with capacity-building, and sharing in the use of the same results data. Because indigenous support for performance management approaches may be limited in many partner countries, donor co-operation strategies will be needed to help build demand and capacity for this aspect of "good governance".

The current trend in donor assistance away from separate projects and towards more collaborative or joint sector assistance approaches, calls for greater donor and partner co-ordination in developing appropriate performance monitoring systems that track higher-order results, particularly at the development objective (sector impact) level. Old methodological tools, such as the project logframe, will need to be replaced by broader frameworks that meet these needs for monitoring results of joint sector programs. Donors need to begin to support and use shared performance monitoring systems owned and maintained by partner countries, rather than creating separate and duplicative data collection systems – even though the indigenous systems might be initially weak and may not necessarily address every specific information need a donor might have.

An obstacle to this vision is the specific requirements for external reporting to domestic stakeholders that many donor agencies now face, limiting their flexibility to agree to and adopt a more country-focused approach shared by other donors and partners. Harmonization may require that donor agencies educate their own domestic stakeholders (e.g., oversight agencies, auditors, legislative branches, etc.), who have often dictated approaches (e.g., terminology, data requirements, analytical methods, reporting formats, etc.) and have demanded that the donor agency's specific contributions to development results be demonstrated. Key stakeholder groups would need to be convinced about the advantages of a collaborative strategy of supporting and using indigenous, country-owned performance monitoring systems, even though this might involve accepting some data quality limitations and less concern over attribution. For example, convincing stakeholders to accept concepts such as "shared accountability" for achieving higher-order development goals would be a major step forward. A united front among donor agencies in support of collaborative principles might assist in convincing sceptical domestic stakeholder groups.

Preliminary lessons learned

Some preliminary lessons learned about establishing effective results based management systems:

- *Allow sufficient time and resources to build effective results based management systems.* Experience shows that it may take up to five or ten years to fully establish and implement performance measurement and management systems. It takes time to develop strategic plans, to monitor results data long enough to establish trends and judge performance vis-à-vis targets, and to evolve new agency decision-making and reporting processes in which performance data are used. Moreover, establishing these new systems appears to be quite costly and labor-intensive. The assumption sometimes made that these new processes can be implemented without additional costs is highly suspect. Without allocating additional funds, it is more than likely that new performance measurement/management activities will compete with traditional evaluation efforts, interfere with implementation activities, or result in the collection of poor quality performance data of questionable use. Either clarify which old procedures and requirements can be dropped or provide adequate additional resources.

- *Keep the performance measurement system relatively simple and user-friendly.* Emphasis should be on keeping the systems simple and management-useful, particularly at the operational level. There is danger that they can become too complex, costly, and time-consuming. In USAID, for example, operating units and implementing partners are beginning to complain that there is no time left for implementing programs, and that much of the higher-order results data collection is not considered directly relevant or useful to them, but is only being used to "report upward". Large numbers of indicators and data can become cumbersome and expensive to collect, maintain, and analyze. The World Bank has noted that excessive concern over indicator/data validity may limit its practical utility as a motivational and management tool. Avoid creating a "measurement bureaucracy" that collects data that are never used.

- *Leadership support for RBM reforms is important.* Without strong advocacy from senior managers, results based management systems are unlikely to be institutionalized broadly or effectively within an agency. Leaders can send strong messages of support for RBM to their staff by giving speeches, sending out agency-wide notices, participating in RBM-oriented workshops, providing adequate budgetary support, etc.

- *Begin with pilot efforts to demonstrate effective RBM practices.* Several donor agencies (including USAID and CIDA) have found it useful to first introduce results based management approaches as pilot efforts in selected country operating units, testing a variety of approaches. After a period of experimentation and lesson learning, the most effective practices were institutionalized agency-wide. Such a gradual approach avoids the dangers of trying to be too comprehensive too fast and also minimizes ineffective practices. Another advantage of this approach is that it builds a base of support for RBM from the "bottom up", as it gains operational legitimacy and relevance.

- *Institutionalize RBM agency-wide by issuing clear guidance.* For effective results based management systems to be established agency-wide, new operational policies and procedures should be spelled out in clear guidance, including statements on roles and responsibilities – who is responsible for what aspects of the systems and procedures.

- *Provide a variety of support mechanisms.* In addition to issuing new policies and procedures, agencies can support their internal organizational efforts to establish effective RBM systems by offering staff "reengineering" training, counselling, technical assistance, supplementary guidance and tools, etc. Some agencies have created central units to "champion" RBM efforts and to provide various types of support to the agency.

- *Monitor both implementation progress and results achievement.* While the current movement to monitoring higher-order results is positive, especially given its historical neglect, this should not be accomplished at the expense of traditional implementation monitoring. Both of these types of monitoring are needed, although for different uses and users. As a project matures, there may be a logical shift from an early emphasis on monitoring implementation to a later focus on monitoring results.

- *Complement performance monitoring with evaluations to ensure appropriate decisions.* Performance monitoring and evaluation should both be viewed as important dimensions of an effective results based management system, that can complement each other nicely. Performance monitoring data alerts managers to performance problems but without further analysis may not present solutions. Experience indicates that performance monitoring data alone are often not adequate for making wise decisions. Evaluations, which examine why performance is good or bad by exploring cause-effect relationships and which typically make action recommendations, are useful complements to simpler presentations of performance monitoring data. Unfortunately, evaluation activity (at least in some agencies) may be on the decline, despite the growing emphasis on results based management. Donor agencies may need to pay more attention to strengthening and integrating the evaluation function within their overall results based management systems, not just in theory but in practice.

- *Ensure the use of performance information, not just for reporting but for management learning and decision-making.* There is growing evidence that these two primary uses of performance information may to some extent be in conflict (e.g., requiring different approaches and methods). Donor agencies need to be aware of these tensions, and attempt to balance their overall RBM systems to accommodate both needs. In particular, they should be on-guard against the possibility that "urgent" demands for performance reporting made by various external stakeholders do not overshadow equally (if not more) important internal management uses of performance information. Too heavy an emphasis on corporate reporting can be disempowering to operational managers and partners.

- *Anticipate and avoid misuses of performance measurement systems.* Experience is accumulating that unless guarded against, performance measurement may be used inappropriately, affecting managers' behaviors in unintended and negative ways, such as leading to program distortions or to dishonest reporting. For example, it may lead to a concentration on those types of activities that are most easily measured rather than on what's most important. Moreover, if resource allocations (or other rewards & penalties) are too rigidly tied to performance data, this may create incentives for biased reporting, as managers try to put their performance in the best possible light. Attempting to hold managers accountable for results beyond their control may also lead to undesirable risk-averse behavior, such as setting targets too low, focusing on outputs rather than outcomes/impacts, avoiding experimental efforts that may have high payoffs because they are risky, etc.

- *Give managers autonomy to manage-for-results as well as accountability.* Managers being held accountable for achieving results should also be empowered with the decision-making authority and flexibility to shift resources away from poorer-performing to higher-performing activities and

projects. Without such authority, managers will be unable to act to improve performance and results, and will soon become sceptical and disillusioned.

- *Build ownership by using participatory processes*. Donor agencies' policies increasingly emphasize participatory approaches involving partners, beneficiaries and stakeholders in all phases of results based management – e.g., in strategic planning exercises, in developing performance measurement systems, and in results-based decision-making processes. Such a participatory approach has the potential to increase the effectiveness of development efforts and builds ownership and commitment of partners and stakeholders to shared objectives. However, it is also likely to be a time-consuming and labor-intensive process.

Next steps for the DAC Working Party on Aid Evaluation

At the last meeting of the DAC Working Party on Aid Evaluation in February 2000, Members strongly supported a proposed second phase of work on results based management by the DAC Secretariat. The second phase will develop a series of good practices notes, drawing not only on this document review (the product of phase one), but also on more recent experience to be gained from interviews at the headquarters of an expanded number of donor agencies, in order to complement and broaden the findings of the phase one desk review.

Phase two will involve interviewing the managers and users of the results based management systems and investigate more deeply actual practices, implementation issues and lessons learned. It will cover an expanded number of donor agencies in addition to the seven included in the desk review. Interviews will be conducted at headquarters of the following donor agencies during November 2000-June 2001: Ministry of Foreign Affairs in Netherlands, Sida, Norad and Ministry of Foreign Affairs in Norway, USAID, BMZ, GTZ and KFW in Germany, AFD in France, DfID, AusAID, CIDA, Danida, UNICEF, UNDP and the World Bank.

Ten effective practices notes will be prepared by October 2001 and presented to the DAC Working Party on Aid Evaluation at the November 2001 Meeting. Topics will especially address "good practices" in areas where more attention is needed.

ANNEXES

Annex 1: Comparison of Project Logframe Terminology Used by Selected Donor Agencies

Terminology	USAID(1)	DFID(2)	DANIDA(3)	SIDA(4)	CIDA(5)	AUSAID(6)	World Bank(7)	UNDP(8)
Goal Impact	Goal	Goal	Development objective, Impact	Sectoral Objective	Goal Impact	Overall Outcomes, Objectives	Goal Impacts	Impacts
Purpose Outcome	Purpose	Purpose	Immediate objective, Outcome, effect	Immediate objectives,	Purpose, Outcome	Intermediate outcome	Purpose, Outcome	Outcome
Outputs	Outputs	Outputs	Outputs	Outputs	Outputs	Outputs	Outputs	Outputs
Activities Processes	Activities	Activities, Processes	Activities, processes	Activities	--	--	Activities, processes	--
Inputs	--	Inputs	Inputs	Inputs	Inputs. Resources	--	Inputs	Inputs

Note: Terminology within a donor agency may vary from source to source. Also, some agencies use different terms to distinguish between hierarchy levels and the type of indicator associated with each level, while other agencies do not. Not all agencies have five levels in their logframe hierarchy. In particular, the activity/process level is often absent.

_____

1. USAID, Genesys, *Gender in M&E, A tool for developing M&E Plans*, August, 1994, p.13. (Note: USAID Automated Directive System (ADS) on M&E refers to inputs, processes and outputs at the "activity" (formerly project) level and to intermediate results and strategic objectives at the country programme level).
2. DFID (formerly (ODA), *A Guide to Appraisal, Design, Monitoring, Management and Impact Assessment of Health and Population Proj*ects, October 1995.
3. DANIDA, *First Guidelines for on Output and Outcome Indicator System*, September 1998, p.7, DANIDA presentation at DAC Workshop on Performance Management and Evaluation, October, 1998.
4. SIDA, *Evaluation Manual for SIDA*, 1994, p.54.
5. CIDA, Presentation at DAC Workshop on Performance Management and Evaluation, October 1998.
6. AUSAID, Presentation at DAC Workshop on Performance Management and Evaluation, October 1998
7. World Bank, *Designing Project Monitoring and Evaluation*, i n Lesson and Practices Series, June 1996, No. 8.
8. UNDP, *Results-Oriented M&E*, 1997, Glossary.

Annex 2.1: DFID's Output and Performance Analysis

Statement of Purpose: DFID's aim is the elimination of poverty in poorer countries.

Objective: Policies and actions which promote sustainable livelihoods

Associated International Development Target:
A reduction by one half in the proportion of people living in extreme poverty by 2015

Other Targets:
% of relevant bilateral projects likely to fully or largely meet their objectives up from 64% to75% by 2002
% of bilateral country programme resources allocated to low income countries increased from 67% to 75% by 2002

Indicators:
% of relevant bilateral projects likely to fully or largely meet their objectives
% of bilateral country programme resources allocated to low income countries
Success in inducing improvement in effectiveness of the international system
Rate of real GDP per capita growth in top UK development partners
Share in GDP of poorest 20% of population in top 30 UK development partners

Objective: Better education, health and opportunities for poor people

Associated International Development Target:
Universal primary education in all countries by 2015
Progress towards gender equality and empowerment of women, by eliminating gender disparity in primary and secondary education by 2005
A two-thirds reduction in the under 5 mortality rate by 2015
A reduction by three-fourths in maternal mortality by 2015
Access to reproductive health services for all by 2015

Other Targets:
% of relevant bilateral projects likely to fully or largely meet their objectives up from 64% to 75%by 2002
(a) Under-5 and (b) maternal mortality in top 30 UK development partners down from 74 to 70 per 1000 live births and from 324 to 240 per 100,000 live births respectively by 2002
% of children in primary education up from 61% to 91% in top 30 UK development partners by 2002

Indicators:
% of relevant bilateral projects likely to fully or largely meet their objectives
(a) under-5 and (b) maternal mortality in top 30 UK development partners
% children in primary education in top 30 UK development partners
Gender disparity in secondary education in top 30 UK development partners
% of population with access to safe water in top 30 UK development partners

Objective: Protection and better management of the natural and physical environment

Associated International Development Targets:
Implementation of national strategies for sustainable development in all countries by 2005
Reversal in current trends in environmental resource loss at global and national levels by 2015

Other Targets:
% of relevant bilateral projects likely to fully or largely meet their objectives up from 64% to 75% by 2002
Agreement of national strategies for sustainable development in top 30 UK development partners

Indicators:
% of relevant bilateral projects likely to fully or largely meet their objectives
national strategies for sustainable development in top 30 UK development partners

Departmental Operations

Targets and Indicators:
Development of a departmental efficiency indicator
Reduction in central and support costs as % of aid policy and administration
Introduction and testing of a new system for scoring on-going projects
Reduction in sickness absence rates
Timely payment of undisputed bills
Replies to Ministerial correspondence within 21 days of receipt
Regular and systematic review of services and their delivery (Better Quality Services Initiative)
Availability of DFID Intranet
Development of performance reporting information system (PRISM)
Increased awareness of fraud
Introduction of procurement cards for UK administrative expenditure
Strengthening of local procurement capacities in overseas offices
Progress towards meeting Investors in People Initiative
Public knowledge of and attitudes to development issues

_____

Source: DFID, *Output and Performance Analysis and notes; An Approach to Portfolio Review in DFID,* 1999.

Annex 2.2: USAID's Strategic Framework

USAID's Mission: Enduring economic and social progress achieved.

Goal: Broad-based economic growth and agricultural development encouraged

Objectives:
Critical private markets expanded and strengthened
More rapid and enhanced agricultural development and food security encouraged
Access to economic opportunity for the rural and urban poor expanded and made more equitable

Targets:
Average annual growth rates in real per capita income above 1 percent achieved
Average annual growth in agriculture at least as high as population growth achieved in low income countries
Proportion of the population in poverty reduced by 25 percent
Openness and greater reliance on private markets increased
Reliance on concessional foreign aid decreased in advanced countries

Indicators:
GNP per capita average annual growth rate (in constant prices)
Difference between average annual growth rate of agriculture and average annual growth rate of population
% of population below poverty line
Trade of goods and services average annual growth rate
Foreign direct investment average annual growth rate
Economic Freedom Index
Aid as % of GNP

Goal: Democracy and good governance strengthened

Objectives:
Rule of law and respect for human rights of women as well as men strengthened
Credible and competitive political processes encouraged
The development of politically active civil society promoted
More transparent and accountable government institutions encouraged

Targets:
Level of freedom and participation improved
Civil liberties and/or political rights improved

Indicators:
Number of countries classified by Freedom House as free/partly free/not free
Freedom House scored for political rights
Freedom House scores for civil liberties

Goal: Human capacity built through education and training

Objectives:
Access to quality basic education, especially for girls and women, expanded
The contribution institutions of higher education make to sustainable development increased

Targets:
% of primary school age population not enrolled reduced by 50%
Difference between girls' and boys' primary enrolment ratio virtually eliminated
Primary school completion rates improved
Higher education increased by 100%

Indicators:
Net primary enrolment ratio
Gross primary enrolment ratio
Ratio of girls' enrolment to boys' enrolment ratio
% of cohort reaching grade 5
% of relevant age group enrolled in tertiary education

Goal: World population stabilized and human health protected

Objectives:
Unintended and mistimed pregnancies reduced
Infant and child health and nutrition improved and infant and child mortality reduced
Deaths, nutrition insecurity, and adverse health outcomes to women as a result of pregnancy and child birth reduced
HIV/AIDS pandemic in developing countries reduced
The treat of infectious diseases of major public health importance reduced

Targets:
Fertility rate reduced by 20 percent
Mortality rates for infants and children under the age of five reduced by 25%
Maternal mortality ratio reduced by 10%
Rate of increase of new HIV infections slowed
% of underweight children under 5 reduced

Indicators:
Total fertility rate
Under 5 mortality rate
% of underweight children under 5
Early neonatal mortality rate (proxy for maternal mortality rate
HIV seroprevalence rate in 15-49 year-olds

Goal: The world's environment protected for long-term sustainability

Objectives:
The threat of global change reduced
Biological diversity conserved
Sustainable urbanisation including pollution management promoted
Use of environmentally sound energy services increased
Sustainable management of natural resources increased

Targets:
National environmental management strategies prepared
Conservation of biologically significant habitat improved
Rate of growth of net emissions of greenhouse gases slowed
Urban population's access to adequate environmental services increased
Energy conserved through increased efficiency and reliance on renewable sources
Loss of forest area slowed

Indicators:
National environmental management strategies
Nationally protected areas
Carbon dioxide emissions
% of urban population with access to safe water
% of urban population with access to sanitation services
GDP per unit of energy use
% of energy production from renewable resources
Annual change in total, natural and plantation forest area

Goal: Lives saved, suffering associated with natural or man-made disasters reduced, and conditions necessary for political and/or economic development re-established

Objectives:
The potential impact of crises reduced
Urgent needs in times of crisis met
Personal security and basic institutions to meet critical intermediate needs and protect human rights re-established

Targets:
Crude mortality rate for refugee populations returned to normal range within 6 months of onset of emergency situation
Nutritional status of children under 5 and under populations made vulnerable by emergencies maintained or improved
Conditional for social and economic development in post-conflict situations improved
Freedom of movement, expression and assembly and economic freedoms in post-conflict situations increased

Indicators:
Crude mortality rate in emergency situations
% of children under 59 months in emergency situations who are wasted
Number of people displaced by open conflict
Changes in number and classification of post-conflict countries classified by Freedom House as free/partly free/not free
Economic Freedom Composite Index

Annex 2.2: USAID's Strategic Framework (continued)

Goal: USAID remains a premier bilateral development agency

Objectives:
Responsive assistance mechanisms developed
Programme effectiveness improved
U.S. commitment to sustainable development assured
Technical and managerial capacities of USAID expanded

Targets:
Time to deploy effective development and disaster relief resources overseas reduced
Level of USAID-managed development assistance channelled through strengthened U.S.-based and local non-governmental organisations increased
Contacts and co-operation between USAID's policy and programme functions and those of other U.S. government foreign affairs agencies expanded
The OECD agenda of agreed development priorities expanded
Capacity to report results and allocated resources on the basis of performance improved

Indicators:
% of critical positions vacant
% of USAID-managed development assistance overseen by U.S. and local private voluntary organisations
Statements at the objective level across strategic plans of U.S. executive agencies concerned with sustainable development are consistent
Number of jointly defined OECD development priorities
Financial and programme results information readily available
Time to procure development services reduced

_____

Source: USAID *Strategic Plan*, September 1997
Note: USAID targets (or performance goals) are 10 year targets, and are for the most part based on the international development goals/targets developed under the *Shaping the 21st Century* initiative of the DAC.

Annex 2.3: The World Bank's Scorecard

Tier 1.A: Development Outcomes
(based primarily on international development goals/indicators)

Development Outcome: Poverty reduction

Indicators:
% of population below $1 per day
Malnutrition - prevalence of underweight below age 5

Development outcome: Equitable Income Growth

Indicators:
Per capita GNP
% share of poorest fifth in national consumption

Development outcome: Human development

Indicators:
Net primary enrolment
Under 5 mortality rate
Ratio of girls/boys in primary education
Ratio of girls/boys in secondary education

Development outcome: Environmental sustainability

Indicators:
Access to safe water
Nationally protected areas
Carbon dioxide emissions

Tier 1.B: Intermediate Outcomes

Intermediate outcome: Policy Reform

Indicators:
Being defined

Intermediate outcome: Institutional Capacity

Indicators:
Being defined
Intermediate outcome: Resource Mobilisation

Indicators:
Being defined

Tier 2: Strategy Effectiveness

A. Impact of Country Strategies

Indicators:
Achievement of Bank Group progress indicators in CAS matrix
OED CAR/CAN ratings on relevance, efficacy, and efficiency
Client/partner feedback

B. Impact of Sector Strategies

Indicators:
Achievement of Bank Group progress indicators in SSP matrix
OED sector study ratings
Client/partner feedback

Tier 3: Process and Capacity

A. Responsiveness, Collaboration, and Partnership

Responsiveness

Indicators:
Lending service standards
Portfolio service standards
Non-lending service (NLS) efficiency

Collaboration

Indicator:
Client satisfaction - survey results

Partnership

Indicators:
Consultative group meetings in the field
Country focused partnership frameworks with development partners
Resources mobilized to client countries through partnerships

B. Human and Intellectual Capital

Product Innovation

Indicators:
Climate for innovation - focus groups
Intensity of innovation - innovative proposals for support and their mainstreaming

Knowledge Management

Indicators:
System coverage regions
System coverage networks

Human Resources

Indicators:
Staff skills index – skills management
Diversity index
Work climate index
Stress indicator

C. Strategies

CAS

Indicators:
Design of country strategies (CASs)
Implementation of CASs

SSP

Indicators:
Design of sector strategies (SSPs)
Implementation of SSPs

D. Products

Deliverable volume

Indicators:
NLS Volume - $m budget and number
Lending approvals - $bn committed and number
Disbursements – gross and net ($bn)

Product Quality

Indicators:
Quality of economic and sector work
Quality at entry
Quality of supervision
Proactivity index
% of projects rated satisfactory at completion by OED

E. Financial and Cost Performance

Productivity

Indicators:
Productivity index
Front-line services as % of net administrative costs

Financial Performance (IBRD)

Indicators:
Net income ($m)
Income as % of administrative expense
_____

Source: World Bank, *Performance Management in the World Bank,* paper presented to the DAC Workshop on Performance Management and Evaluation, October 1998

Note: Targets have not yet been determined for the Scorecard indicators. Some indicator definitions are yet to be defined. The development objectives and indicators of Tier 1.a are largely the international development goals and indicators developed under the *Shaping the 21st Century* initiative of the DAC.

Annex 2.4: AusAID's Performance Information Framework

AusAID's outcome: Australia's national interest advanced by assistance to developing countries to reduce poverty and achieve sustainable development

KRA: Improve agricultural and regional development in developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

KRA: Increase access and quality of education in developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

KRA: Promote effective governance in developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

KRA: Improve health of people in developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

KRA: Provide essential infrastructure for people in developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

KRA: Deliver humanitarian and emergency assistance to developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

KRA: Promote environmental sustainability in developing countries

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
      a)   Number of people assisted
      b)   Number and type of outputs (services and goods) provided

Annex 2.4: AusAID's Performance Information Framework (continued)

KRA: Promote equal opportunities for men and women as participants and beneficiaries of development

Target:
75% of projects receive a quality rating of satisfactory overall or higher

Indicators:
Expenditure, $m (cost indicator)
Number of projects implemented (quantity indicator)
% of projects receiving a quality rating of satisfactory overall or higher (quality indicator)
Significant project outputs achieved, e.g.
        a)   Number of people assisted
        b)   Number and type of outputs (services and goods) provided

_____

Source: AusAID, Attachment A: Performance Information Framework

---

Annex 2.5: The UNDP's Strategic Results Framework

Goal: Promote decentralisation that supports participatory local governance, strengthens local organisations and empowers communities

Results Indicators:
Programme outcome and output indicators: selected and reported by country operating units within specific strategic areas of support

Situational indicators:
Selected by UNDP headquarters and reported by country operating units

Goal: Promote poverty focused development

Results Indicators:
Programme outcome and output indicators: selected and reported by country operating units within specific strategic areas of support

Situational indicators:
Selected by UNDP headquarters and reported by country operating units

Goal: Equal participation and gender equality concerns in governance and economic and political decision-making at all levels

Results Indicators:
Programme outcome and output indicators: selected and reported by country operating units within specific strategic areas of support

Situational indicators:
Selected by UNDP headquarters and reported by country operating units

Goal: Promote integration of sound environmental management with national development policies and programmes

Results Indicators:
Programme outcome and output indicators: selected and reported by country operating units within specific strategic areas of support

Situational indicators:
Selected by UNDP headquarters and reported by country operating units

Goal: Special development situations (crisis countries)
Goal: UNDP support to the UNDP
Goal: Management

Source: UNDP

Annex 2.6: Danida's Output and Outcome Indicator System

Overall Goal: Poverty Reduction

Assistance sector: Agriculture

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example -- for sub-sector: improving farmers access to credit:

Number of farmers having formal access to credit
Number of farmers having formal credit through Danish assistance
Number of these farmers having or having had a loan
Number of these farmers who are women

Assistance sector: Education

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example • sub-sector: Education access

Net enrolment rate in primary education
Girls enrolled as % of total net enrolment rate
Total retention rate
Dropout rate for girls

Assistance sector: Environment

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example -- sub-sector: capacity-building, government

Number of staff trained on-the-job
Number of staff trained on certified or tailor-made courses
Number of organisations targeted for training

Assistance sector: Good governance

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example -- sub-sector: legal aid:

Number of legal aid clinics established
Number of female professionals in these clinics
Number of persons assisted
Number assisted who were women

Assistance sector: Health

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example • sub-sector: preventative health interventions

Children immunization by age 12 months- measles
Pregnant women- tetanus toxoid

Assistance sector: Infrastructure (transport, electrification, telephones)

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example -- sub-sector: roads improvement

Kilometre's reconstructed
Kilometres rehabilitated
Kilometres repaired
Increase in traffic as result of improvements
Length of main road improved through Danida funding as % total road network
Length of main road improved through Danida funding as % total need for improvement

Annex 2.6: Danida's Output and Outcome Indicator System (continued)

Assistance sector: Water (water resources, drinking water, sanitation)

Selected national indicators:
Selected by Danida headquarters and reported by country operating units

Results indicators:
Standard project/programme output and outcome indicators selected by headquarters and reported by country operating units within specific sub-sectors

Example -- sub-sector: access to clean water in rural areas

Number of hand pumps installed/rehabilitated
Number of persons served by these hand pumps
Number of spring protections and wells constructed/rehabilitated
Number of persons served by these spring protections or wells
Number of pipe schemes constructed
Number of persons served per scheme
Total number of water points installed/rehabilitated
Total number of persons served
Total number of persons served by Danida financed projects/components as contribution to national estimated average

_____

Source: Danida, *First Guidelines for an Output and Outcome Indicator System*, September 1998

SELECTED REFERENCES

AusAID

AusAID, *Development of Performance Information in the Australian Agency for International Development*, paper prepared for the DAC Workshop on Performance Management and Evaluation, New York, October 1998.

Parsons, Stephen A., *AusAID's Activity Management System and the Development of Performance Information Systems*, October 1998.

AusAID, *Enhancing Quality in the Aid Program*, slide presentation.

AusAID Circular No 7 of 1999, *Performance Information Under Accrual Budgeting.*

AusAID Circular No 20 of 30 June 1999, *Activity Monitoring Brief Operational Guidelines*.

CIDA

CIDA, *Results-Based Management in CIDA: An Introductory Guide to the Concepts and Principles*, January 1999.

CIDA, *Guide to Project Performance Reporting: For Canadian Partners and Executing Agencies*, May 1999.

CIDA, *Measuring Performance at the Development Agency Level*, Workshop Proceedings, May 1996.

CIDA, *Performance Newsletter: Focus on RBM Implementation*, Vol. 2 No. 1, March 1999.

CIDA, *Performance Management at CIDA*, presentation materials and slides for the DAC Workshop on Performance Management and Evaluation, New York, October 1998.

CIDA, *Policy Statement on Results-Based Management in CIDA*, 1996.

CIDA, *Framework of Results and Key Success Factors*.

Danida

Danida, *First Guidelines for an Output and Outcome Indicator System*, September 1998.

Jespersen, Lis and Klausen, Anne-Lise, Danida's presentation outline and slides for the DAC Workshop on Performance Management and Evaluation, New York, October 1998.

DFID

DFID, *Presentation by Department for International Development, United Kingdom*, for the DAC Workshop on Performance Management and Evaluation, October 1998.

DFID, *Output and Performance Analysis and notes.*

DFID, *An Approach to Portfolio Review in DFID*, Evaluation Department, June 1999.

UNDP

UNDP, *UNDP Results Framework: Overview*, March 1999.

UNDP, *UNDP Results Framework: Technical Note*, revision 1, March 1999.

UNDP, *Signposts of Development: Selecting Key Results Indicators*, May 1999.

UNDP, *Results Based Management in UNDP: A Work in Progress, Learning and Applying as You Go*, presentation materials for the DAC Workshop on Performance Management and Evaluation, New York, October 1998.

UNDP, *Results-Oriented Monitoring and Evaluation*, Office of Evaluation and Strategic Planning, Handbook Series, 1997.

USAID

USAID, *Trip Report: Workshop on Performance Management and Evaluation*, prepared by Annette Binnendijk, October 1998.

USAID, *Managing for Results at USAID*, prepared for the Workshop on Management and Evaluation, New York, October 5-7, 1998 by Annette Binnendijk.

USAID, *Programme Performance Monitoring and Evaluation at USAID*, by Scott Smith, January 1996.

USAID, *Automated Directives System*, series 200.

USAID, *Performance Monitoring and Evaluation Tips*, Numbers 1-11.

USAID, *Performance and Budget Allocations*, September 1998.

World Bank

Jarvie, Wendy, *Performance Management in the World Bank*, presentation to the DAC Workshop on Performance Management and Evaluation, New York, October 1998.

McAllister, Elizabeth, "Results Based Management", in *OED Views*, Vol 1, No 1, Operations Evaluation Department, World Bank.

Poate, Derck, "Designing Project Monitoring and Evaluation", in *Lessons & Practices*, Operations Evaluation Department, World Bank, No 8, June 1996.

Weaving, Rachel and Thumm, Ulrich, "Evaluating Development Operations: Methods for Judging Outcomes and Impacts", in *Lessons & Practices*, Operations Evaluation Department, world Bank, No 10, November 1997.

Picciotto, Robert, and Rist, Ray (editors), *Evaluating Country Development Policies and Programs: New Approaches for a New Agenda*, in *New Directions for Evaluation*, a publication of the American Evaluation Association, No 67, Fall 1995.

Hanna, Nagy, *1999 Annual Review of Development Effectiveness*, Operations Evaluation Department, World Bank, 1999.

Wolfenson, James D., *A Proposal for a Comprehensive Development Framework*, January 1999. *Comprehensive Development Framework: Report on Country Experience*, World Bank, September 2000.

Other References

Poate, Derek, *Measuring & Managing Results: Lessons for Development Cooperation*, Office of Evaluation and Strategic Planning, UNDP, 1997.

"Measuring and Managing Results: Lessons for Development Cooperation", article *in Sida Evaluations Newsletter*, 6/97.

Helgason, Sigurdur (Public Management Service, OECD*), Performance Management Practices in OECD Countries,* a paper delivered at the DAC Workshop on Performance Management and Evaluation (DAC Working Party on Aid Evaluation, SIDA/UNDP), October 1998.

Cooley, Larry, *The Concept of Performance Management*, presentation materials for the DAC Workshop on Performance Measurement and Evaluation, October 1998.

DAC Working Party on Aid Evaluation*, Workshop on Performance Measurement and Evaluation*, Room Document No. 5 for the 30th Meeting, submitted by Sweden, May 1998.

DAC Working Party on Aid Evaluation, *Rating Systems in Aid Management: Executive Summary*, Note by the Delegation of the Netherlands, the United Kingdom and the Secretariat, for the Meeting in October 1996.

DAC Working Party on Aid Evaluation*, Review of Current Terminology in Evaluation and Results Based Management*, December 1999.

DAC Working Party on Aid Evaluation, *Glossary of Terms in Evaluation and Results Based Management,* Annexes, Background Document No. 3 for the Meeting in February 2000.

OECD/DAC, *Shaping the 21st Century: The Contribution of Development Co-operation*, May 1996.

OECD, *In Search of Results: Public Management Practices*, 1997.

OECD Public Management Service, *Best Practice Guidelines for Evaluation*, 1998.

OECD, *Budgeting for Results: Perspectives on Public Expenditure Management*, 1995.

United States General Accounting Office (GAO), *Programme Evaluation: Agencies Challenged by New Demand for Information on Programme Results,* (GAO/GGD-98-53), April 1998.

GAO, *Managing For Results: Analytic Challenges in Measuring Performance*, (GAO/HEHS/GGD-97-138), May 1997.

GAO, *Past Initiatives Offer Insights for GPRA Implementation*, (GAO/AIMD-97-46, March 1997.

GAO, *Performance Budgeting: Initial Agency Experiences Provide a Foundation to Assess Future Directions*, (GAO/T-AIMD/GGD-99-216), July 1999.

GAO, *Executive Guide: Effectively Implementing the Government Performance and Results Act*, (GAO/GGD-96-118), June 1996.

GAO, *Performance Measurement and Evaluation: Definitions and Relationships* (Glossary), prepared by Susan S. Westin and Joseph Wholey, (GAO/GGD-98-26), April 1998.