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

EXECUTIVE SUMMARY

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

The Development Assistance Committee (DAC) Working Party on Aid Evaluation is an international forum where bilateral and multilateral development evaluation experts meet periodically to share experience to improve evaluation practice and strengthen its use as an instrument for development co-operation policy.

It operates under the aegis of the DAC and presently consists of 30 representatives from OECD Member countries and multilateral development agencies (Australia, Austria, Belgium, Canada, Denmark, European Commission, Finland, France, Greece, Ireland, Italy, Gernamy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States, World Bank, Asian Development Bank, African Development Bank, Inter-American Development Bank, European Bank for Reconstruction and Development, UN Development Programme, International Monetary Fund, plus two non-DAC Observers, Mexico and Korea).

Further information may be obtained from Hans Lundgren, Advisor on Aid Effectiveness, OECD, Development Cooperation Directorate, 2 rue André Pascal, 75775 Paris Cedex 16, France. Website: http://www.oecd.org/dac/evaluation.

EXECUTIVE SUMMARY

1. Public sector reforms in the OECD countries

During the 1990s, many of the OECD countries have undertaken extensive public sector reforms in response to economic, social and political pressures. Budget deficits, structural problems, growing competitiveness and globalization, lack of public confidence in government, and growing demands for better and more responsive services and for more accountability have all been contributing factors. Often, government-wide legislation or executive orders have driven and guided the public sector reforms.

While there have been variations in the reform packages implemented in the OECD countries, there have been many common aspects. For example: the focus on performance issues and on achieving results; the devolution of management authority and responsibility; an orientation to customer needs and preferences; an emphasis on participation and teamwork; the reform of budget processes and financial management systems; and the application of modern management practices.

2. Results based management

A central feature of the reforms has been the emphasis on improving performance; that is, on ensuring that government activities achieve desired results. 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.

A key component of results based management is performance measurement, which is the process of objectively measuring how well an agency is meeting its stated goals or objectives. 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 and reporting those results vis-à-vis the targets. While performance measurement is concerned more narrowly with the production *or supply* of performance data, performance management is broader. It is equally concerned with generating management *demand* for performance information -- that is, with its uses in management decision-making processes and with establishing various organizational 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.

3. Results based management in the donor agencies

As has been the case more broadly in the OECD country public sectors, 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 the government-wide reforms have all contributed to the donor agencies' recent efforts to establish results based management systems.

This paper focuses on the experiences and approaches of the donor agencies with establishing results based management systems. It is based on a document review of selected donor agencies with the most experience, including five bilateral and two multilateral agencies: USAID (United States); DFID (United Kingdom); AusAID (Australia); CIDA (Canada); Danida (Denmark); the UNDP; and the World Bank.

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. There is less documented experience with the actual use of performance information for internal management decision-making.

Results based management and measurement processes take place at three key organizational 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 a number of the donor agencies to establish country level systems, usually implemented by 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 urgent importance in many donor agencies as they face increasing public pressures and government-wide mandates requiring annual reporting on agency-wide performance and results.

4. Phases of results based management

Key elements or phases of results based management include:

i. Identifying clear and measurable objectives (results), aided by logical frameworks.
ii. Selecting indicators that will be used to measure progress towards each objective.
iii. Setting explicit targets for each indicator, used to judge performance.
iv. Developing performance monitoring systems to regularly collect data on actual results.
v. Reviewing, analyzing and reporting actual results vis-à-vis the targets.
vi. Integrating evaluations to provide complementary performance information not readily available from performance monitoring systems.
vii. Using performance information for internal management accountability, learning and decision-making processes, and also for external performance reporting to stakeholders and partners.

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

5. Other components of results based management

Other reforms are often associated with results based management systems. Frequently, these other components act to reinforce or facilitate the use of performance information. Some of these organizational changes include:

- Accountability -- instituting new mechanisms for holding agency managers and units accountable for achieving results at appropriate levels.
- Decentralization -- delegating authority to the management level accountable for results, and empowering them with flexibility to shift resources to better performing activities.

- Client focus -- consulting with beneficiary groups concerning their preferences and satisfaction with goods and services provided, and being responsive to their needs.
- Participation and partnership -- involving partners and stakeholders in all aspects of performance measurement and management processes, and seeking greater harmonization of efforts.
- Reformed operational policies and procedures -- instituting new policy and procedural directives aimed at changing the way the agency conducts its business.
- Supportive mechanisms -- assisting managers to effectively implement performance measurement and management in various ways, such as providing training, technical assistance, performance information databases, guidebooks, tips and best practices series.
- Cultural change -- equally important for successful results based management is transforming the organizational culture -- that is, the values, attitudes and behaviors of its personnel.

6. Special challenges facing the donor agencies

Because of the nature of development co-operation work, the donor agencies face special challenges and issues 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. This can make establishing performance measurement systems more complex and costly than normal. For example, development 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's little performance measurement experience.
- Often lack standard indicators on results that can be easily aggregated across projects.
- Are usually only minor actors affecting impacts, with consequent problems in attributing them to their agency's activities.
- Typically rely on outcome and impact data collected by partner countries, who have limited technical capacity and resources, with consequent quality, coverage and timeliness problems.

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

7. Performance measurement at the project level

Performance measurement at the project level is concerned with measuring both a project's implementation progress and results achieved. i) *Implementation measurement* is concerned with whether or not project inputs and activities are in compliance with design budgets, workplans, and schedules, whereas ii) r*esults measurement* is concerned with whether or not actual results are achieved as planned. Results are usually measured at three levels – immediate outputs, intermediate outcomes and long-term impacts. While 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)

Measuring performance at the project level can be divided into five phases, as briefly outlined below. The donor agencies stress the importance of participatory or collaborative approaches in all phases of performance measurement – that is, involving not only donor agency project managers but also representatives from the implementing agency, the partner government, the intended beneficiary groups, and other stakeholders. This helps build agreement around the project's objectives and commitment to the performance measurement process.

7.1 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 logframe is a favorite tool used by development agencies for conceptualizing a project's objectives and the strategies that will be used to attain them. The logframe is typically based on a five-level hierarchy model with logical cause-effect relationships among them, with those at the lower level of the hierarchy contributing to the attainment of those above. Thus, *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 *project goal*.

7.2 Selecting indicators

Next, performance indicators are selected for measuring progress in implementing activities and in achieving results. The logframe provides a multi-level structure around which the indicators are typically constructed. Indicators specify what to measure along a scale or dimension in order to gauge progress (e.g., number of workshops held, percentage of farmers attending demonstration sessions, crop yields, etc.). The relative importance of indicator types is likely to change over the project's life cycle, with more emphasis on input and process indicators at first, then shifting to output, outcome (purpose-level), and impact (goal-level) indicators later on as the project matures. Also, different management levels tend to place emphasis on different indicator types. For example, project field staff will find input and process indicators of most use, whereas project managers will be more interested in achievement of project outputs and outcomes. Senior agency officials will be interested in the longer-term and broader social and economic impacts of the project, which may not be evident until after the project is completed. These different intended uses and users need to be kept in mind when selecting indicators.

Many of the donor agencies have devised checklists of criteria against which proposed indicators can be judged and selected. For example, some commonly used criteria include:

- Valid -- Does the indicator directly represent the result it is intended to measure?
- Objective -- Is the definition precise and unambiguous about what is to be measured?
- Reliable -- Is the data consistent or comparable over time?
- Practical -- Can data be collected easily, on a timely basis and at reasonable cost?
- Useful -- Will the data have utility for decision-making and learning?
- Owned -- Do partners and stakeholders agree that this indicator makes sense to use?

Tradeoffs among these indicator selection criteria may exist. Probably the most important, overarching consideration is that the indicators provide managers with the information they need to do their job. While indicator data should be of sufficient quality to be credible and ensure the right decisions are made, it also needs to be practical - timely and affordable. The number of indicators selected should be limited to the minimum needed to adequately capture the key dimensions of a result. Keeping the performance measurement system simple will avoid overburdening managers and staff with unnecessary data collection responsibilities.

7.3 Setting targets

Once indicators have been identified, actual baseline values should be collected for each indicator, before the project activities get underway. This will be important for gauging whether progress is being made later. Often agencies also set explicit targets - that is, a particular value for an indicator to be accomplished within a given timeframe (e.g. contraceptive prevalence rate will be increased to 65% 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.

A natural tension exists between setting targets that are high enough to make project managers and staff stretch to achieve them, and yet low enough to be realistic and achievable. If they are set unrealistically high and unattainable, confidence and credibility will suffer and may even set in motion perverse incentives to hide or distort the figures. Any information that helps to ground a target setting exercise and ensure its realism is useful. For example, it is useful to establish a baseline, identify historical trends, seek implementing agency staff and customer views, survey expert opinion about what is possible, review research findings, or identify benchmarks (i.e., compare what results have been achieved by similar projects with a reputation for high performance).

7.4 Monitoring performance (collecting data)

Once indicators have been selected, baselines established and targets set, actual data for each indicator is collected at regular intervals. It may be useful to distinguish between two types of performance monitoring:

- Implementation monitoring involves the frequent, 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 budget.
- 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. This type of monitoring demonstrates whether a project is moving towards its objectives.

Project managers have found it useful to prepare performance monitoring plans to record key aspects of data collection, such as providing definitions for each indicator, source and methods of data collection, frequency/schedule for collection, and assignment of responsibility for collection.

Some common patterns in data collection approaches can be observed to vary according to different levels of the project logframe hierarchy. These common patterns include typical variations in data collection sources/methods, frequency of collection, and assignment of responsibility. As one moves to higher and higher levels of the logframe hierarchy, there is the tendency for data collection efforts to become more expensive, time-consuming, and technically complex. Also, there is a tendency for data collection efforts to be conducted less frequently. The placement of responsibility for data collection also tends to shift from the implementing agency at the lower levels to the donor agency and/or to the partner government at the higher levels.

Data on project inputs, processes, and outputs are generated mostly by project staff and are based on simple reporting systems updated frequently. Data on outcomes are generally collected periodically (e.g., annually) 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. Impact data are usually only collected every few years or at the project’s beginning and end (or ex post). 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 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.

At what level should the focus of performance monitoring be placed? Concentrating on just one level of the logframe hierarchy may have unintended, even dysfunctional, consequences. For example, concentrating only on the output level may result in "doing the wrong things well". Concentrating only on higher outcome and impact levels 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 possible, within practical limits. Developing a more comprehensive performance monitoring system that recognizes 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. On the other hand, there may be some tradeoffs between developing a system that is comprehensive, versus the need to keep the system relatively simple and avoid overburdening capacity and drawing resources away from other necessary functions.

7.5 Reviewing and reporting performance data

Project performance monitoring data is periodically reviewed and reported by project management. These reviews typically involve a simple analysis of project effectiveness in achieving its targets, by comparing actual results with planned results. However, analysis of performance monitoring data may address a broad variety of issues. For example:

- Economy -- the relationship between costs and physical inputs.
- Efficiency -- the relationship between costs and outputs.
- Productivity -- relationships between inputs and outputs.
- Excellence/quality - producing high quality outputs.
- Equity - extent to which disadvantaged sub-populations have equitable access to results.
- Customer satisfaction - how well project outputs correspond to client preferences.
- Effectiveness - extent to which results (outputs, outcomes, impacts) are achieved as planned.
- Attribution - extent to which outcomes and impacts can be attributed to project outputs.
- Cost-effectiveness - relationship between costs and outcomes/impacts attributable to a 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, the partner country’s national development priorities, and to the donor agency’s corporate goals.

Routine performance monitoring and analysis alone may not be adequate for addressing some of the performance issues listed above (e.g., attribution, cost-effectiveness, sustainability), which because of their long-term nature and/or complexity may require special in-depth evaluation studies.

A number of donor agencies have established performance rating systems whereby project managers, drawing on data from performance monitoring systems, judge their own project’s performance. This is done by assigning a rating along a scale (e.g., highly satisfactory, satisfactory, unsatisfactory, or highly unsatisfactory) against a number of performance criteria (e.g., effectiveness, efficiency, relevance, sustainability, etc.). These performance ratings are typically reported to agency headquarters in standard reporting formats at specific times, such as at project completion or in annual progress reports. In some agencies, these self-assessments are independently reviewed or audited to help ensure their credibility. A particularly useful characteristic of project performance rating systems is that they enable consistent comparisons and aggregation across the project portfolio.

Focusing on the project as the unit of analysis for performance measurement has increasingly come under criticism. Limitations of the project logframe include its lack of strategic and long-term focus, and its inappropriateness for dealing with newly emerging country sector-wide programming modes that are jointly supported by numerous development partners.

8. Performance measurement at the country level

A number of donor agencies have recently developed performance measurement systems for broader country programs – usually defined as sets of related projects or non-project activities sharing the same development objective within a partner country. 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 a sector, sub-sector, or a crosscutting objective. Thus, the unit of analysis is not a single project but a whole country program that typically includes many activities implemented by different donor agencies and partner organizations over a relatively long time period.

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 adopts a similar country program model. The World Bank has also recently initiated a more strategic approach to planning, implementing, and monitoring development efforts at the country level, as part of its 1999 Comprehensive Development Framework initiative. This Framework calls on all development partners to make fundamental changes in the way their development operations are conducted at the country level, based on principles of results-orientation, country ownership and participation, partnership, and a shared long-term, holistic vision of development goals.

A conceptual 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 developing country. The results framework relies on objective tree concepts, and diagrams the logical cause-effect relationships between various partners’ activity outputs at the bottom, intermediate results or outcomes in the middle, and the strategic development objective at the top. Thus, it embodies the development hypotheses underlying multiple partners' harmonized strategies for achieving a shared development objective. Results frameworks should be developed via collaborative processes, ideally under the leadership of the partner country government, and involving all donor agencies and other development partners (including civil society and private sector) that are working towards a shared development objective.

Results frameworks are useful as strategic planning and management tools. They help identify what program strategies are necessary and sufficient to achieve a significant development objective, and then enable collaborating partners, working in harmony, to sort out their individual responsibilities or contributions to the overall strategy. This can help a donor's operating units to better focus and concentrate their country assistance activities into a 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 an operating unit usually have to be in line with the donor agency's overall corporate goals and areas of comparative advantage. They should also, of course, be aligned with the partner country's development goals and reflect a rational division of labor among various partners.

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 various partners (including donor operating units) are responsible. Performance data from the monitoring system is used to alert managers when actual results are not meeting targets as planned, indicating the need for corrective adjustments to be made in relevant projects or other activities. It may be useful to occasionally supplement more routine reviews of the performance monitoring data with complementary strategic, program-wide evaluations that assess the relative effectiveness of various strategies and activities for achieving the development objective. A number of donors are increasingly conducting such strategic evaluations; for example the World Bank's country assistance evaluations and country sector impact evaluations. Conducting such country sector program evaluations jointly with other partners has obvious advantages, and is being piloted by members of the DAC Working Party on Aid Evaluation.

Whereas the project approach puts equal weight on monitoring all elements of the logframe hierarchy, and may even traditionally have favored implementation monitoring, the country program results framework puts the higher-order development objective and intermediate outcomes at center-stage. It is less concerned with defining the individual project means (inputs/processes) and outputs, and much more concerned with measuring and achieving the higher-level results. The shift from individual projects to broader programs also implies a different time-frame dimension, freed from the confines of a single project's life cycle. By focusing on country level development objectives and intermediate outcomes, the timeframe now becomes longer-term, outliving the comings and goings of individual project activities.

Individual project activities tend to be less well defined in this approach, allowing for more flexible designs and experimentation, rather than rigid "blueprint" approaches. Moreover, in some agencies, headquarters no longer approves projects. Instead, authority is delegated to operating units in the field so they can shift course mid-stream if results monitoring information indicates certain activities are not working well.

A potential danger with this approach is that by concentrating on the higher-order results, their linkages with project activities may become overly vague or disconnected. Moreover, activity implementation monitoring may be ignored. It is suggested that those adopting the country program approach take time to clearly align their individual project activities into the overall results framework.

This country program level approach puts a premium on partnerships and more collaborative efforts, since achieving a strategic, long-term development objective is clearly dependent on the activities of many development partners -- e.g., the partner countries' government agencies, private sector and civil society organizations, and various donor agencies. Tools such as the country program results framework and related program performance monitoring systems should be particularly well suited to new modes of development co-operation. For example, the emerging medium-term expenditure frameworks and sector-wide programming approaches that are based on country ownership and partnerships focused on achieving shared development objectives.

While this approach holds considerable promise, its actual use has often fallen short of its potential. Donor agency operating units still often develop country program results frameworks and related monitoring systems in relative isolation and from their own agency's perspectives. While there is typically some limited participation by their implementing agency partners and stakeholders, the focus is usually on the agency's own programs and strategies, rather than placing equal focus on all relevant partners' programs. Country-level results frameworks work best when developed jointly, in a collaborative fashion. However, a number of constraints prevent greater participation and joint efforts, including the time- and resource-intensity of co-ordination, differing donor agendas and the concerns over sharing "control", and the difficulty of harmonizing diverse frameworks, methodologies, indicators and terminology being promoted by various donors.

9. Performance measurement at the agency level

Largely driven by domestic public pressures and government-wide legislation for annual reporting on agency performance, the donor agencies are clarifying their overall goals and seeking ways to summarize their achievements vis-à-vis those goals. Measuring and reporting on results at the agency-wide level poses a significant challenge for the donor agencies. As already discussed, they face a number of obstacles in attempting to aggregate results, complicated by the special nature of development co-operation efforts.

All of the donor agencies reviewed have recently developed and issued policy papers or strategic plans that clearly articulate their agency's overall mission and the key development goals or priority areas on which they will concentrate. Most of these corporate-level goals are sector-oriented (e.g. better education, improved health, good governance) although some may be crosscutting thematic concerns (e.g. gender equality, partnerships) or internal management efficiency goals. These statements about agency goals serve to articulate to external audiences what the overall aims of the development assistance program are, and provide a framework or structure for gathering and reporting data on overall agency performance or results achieved. Demonstrating results 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 priority goal areas. Agency country offices in many cases have been asked to align their country development objectives and assistance programs/ projects within the new agency goal structures.

In some cases, agencies have elaborated their key corporate goals into several sub-categories, forming a multi-level framework or hierarchy of objectives. These multi-level strategic frameworks serve to clarify even further what development goals the agency seeks to contribute towards and how it intends to contribute. The hierarchies serve as detailed structures for measuring and reporting on agency results at several levels. For example, a typical three-level hierarchy structure might include agency goals, sub-goals, and supporting program approaches. Some agencies have found it useful to present their strategic frameworks as graphic diagrams or visual displays, using objective tree concepts.

Collecting data for agency-wide performance assessments and reporting takes place from two basic sources; i) from existing international sources (databases) that maintain global and country level statistics on sector development trends, and ii) from the project/program performance measurement systems maintained by the agency's country operating units. Some agencies also integrate 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 or program portfolios. Findings from these agency-wide analyses of aggregate project/program performance and results data are usually reported to stakeholders in annual performance reports.

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 outputs and country-level sector statistics -- aggregation of indicator data may be relatively easy. But in the case of outputs, the question "so what?" may be raised. With country level statistics, it is rarely possible to credibly link changes to a single agency's interventions, especially on a year-to-year basis. In the middle are project outcomes, which should be more significant than outputs, yet should be more clearly attributable to agency activities than national statistical trends. The problem here is that often there is great diversity in project/program outcomes and in their performance measures, so aggregation using comparable standard indicators is often not possible. Some agencies have overcome this by developing rating systems that score a project's (or program's) performance in meeting its intended outcomes and then sum the performance rating across the portfolio – e.g., the numbers and percentages of projects/programs that were successful or unsuccessful in achieving their targets.

Three basic options for aggregating results (including their pros and cons) are discussed below.

I. ***Selecting the project/program 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.

II. ***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. 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).

III. ***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 often 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 than 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/programs 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.

None of these options for aggregating results to the agency-wide level appear to be ideal, and the donor agencies face considerable challenges in their current efforts to summarize and report performance at the corporate level. The donor agencies, for the most part, are still experimenting with different options, and may benefit from mixing or balancing the various approaches available. How to best co-ordinate, synthesize and integrate findings from evaluation reports into annual agency performance reports is another issue needing attention.

10. The role of evaluation vis-à-vis performance measurement

The emergence of new results based management systems has raised the issue of how existing evaluation systems fit in. Performance measurement and evaluation are generally viewed by donor agencies as two distinct functions that provide complementary types (and sources) of performance information, both of which are necessary for effective results based management. However, the definitions and distinctions made between the two functions differ from agency to agency, and in some agencies the differences are not clear-cut. However, some common themes emerge, which are summarized below.

Generally speaking, self-assessments by project or program managers, such as annual performance monitoring reports and completion reports are considered an integral part of the performance measurement and reporting system, while evaluations are viewed as supplementary. Key distinctions that might be drawn between performance measurement/monitoring and evaluation follow:

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

- The performance monitoring reports are typically mandatory for larger projects and thus provide a reasonably complete coverage of the overall project portfolio, whereas evaluations are usually conducted on a much more selective (i.e., occasional, optional) basis for projects or programs of particular interest or concern to management.
- Performance reports involve relatively straightforward presentations of performance data or ratings, following standard, comparable formats that can be easily entered into databases and analyzed 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, 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 or program's context and factors influencing performance.
- Routine performance monitoring can serve as an early warning system to alert managers when there are performance shortfalls. However, they do not assess the causes of the shortfalls nor make recommendations for appropriate management actions, as do evaluations.
- While performance monitoring tends to emphasize quantitative indicators, evaluations often rely on qualitative methods (such as the rapid appraisal techniques), which are better for understanding context and making recommendations.
- Because of timing as well as the need for more in-depth analysis, some performance issues, such as long-term impact, attribution, cost-effectiveness, and sustainability, can probably be better addressed by evaluation than by routine performance monitoring reports.

Thus, evaluations and performance measurement/monitoring can be viewed as distinct but complementary functions. Both are management tools. Both are important sources of performance information that together can contribute to management learning and decision-making processes and to external performance reporting requirements.

While evaluation is increasingly viewed as integral to results based management, some tensions remain. There is some concern within the donor agencies that performance measurement and reporting tasks, often required by government-wide law or executive orders, may be "crowding out" evaluations. That is, they may be competing for the same, increasingly scarce staff time and other resources. Typically, agencies have not been given adequate additional funding and personnel to establish their performance management and measurement systems, while overall agency budgets and manpower have often been on the decline. For example, since the mid-1990s when USAID reengineering reforms mandated the establishment of performance management and measurement systems, the size of USAID's central evaluation office staff and resources has declined considerably. Moreover, the number of evaluations conducted by USAID's country operating units also declined sharply since 1995.

11. 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 now 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 in developing countries. 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.

Since most donor agencies are still in the early stages of establishing their results based management systems, especially at the higher organizational levels, their actual experience with using performance information is based on early impressions and typically is not well documented. Performance information (from both performance measurement and evaluation sources) serves two primary aims or uses in the donor agencies as discussed below.

I. 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.

- *Promote Learning*. Performance information promotes continuous management learning about what results are being achieved by projects/programs and why -- i.e. what factors are influencing good or poor performance. Improved knowledge is a prerequisite for better decisions.
- *Facilitate Decision-making*. Management's learning in turn facilitates their making appropriate decisions. These may include operational-level implementation decisions, or more strategic decisions such as formulating future agency strategies or allocating agency resources. 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 and strategies. A special type of decision-making that performance information is increasingly being called upon to influence is resource allocations across programs and countries.

II. 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 accountability and internal (individual or work unit) accountability. In the case of donor agencies, external accountability involves reporting not only to domestic stakeholders, but extends to development partners. Performance reporting/accountability uses 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. Care needs to be taken to avoid holding units and managers accountable for results that are beyond their control, and to avoid creating perverse incentives (such as distorting performance data, avoiding experimentation and risky projects, etc.).

11.1 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 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. Generally speaking, accountability tends to be backward-looking and often focused on fault-finding, whereas learning/management improvement takes a more positive and forward-looking perspective.

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). Thus, these two aims or uses may work at cross-purposes. How a donor's managerial system resolves these tensions will send strong signals, affecting the organization's cultural values and behaviors. In other words, how performance information is used can strongly influence the behaviors of managers.

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.

These tensions may be resolved to some extent by thinking about accountability in a new way. The old "command and control" view of accountability, holding managers responsible for achieving specific results, could give way to a newly emerging view *of shared accountability* based on partnerships. Under this new paradigm, managers might be held accountable for working with partners to achieve higher-order results, for learning from failures, and for continually using performance information in their decision-making processes. In other words, they would be held accountable for forming partnerships, for learning and for managing-for-results, rather than for achieving specific results, especially those out of their control.

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 potential conflict among performance information uses is more unique to the donor agencies. Donor agencies are accountable not only to domestic stakeholders but also to the partner country stakeholders. To the extent that donor agencies' performance measurement and reporting systems may vary considerably from one to the next, partner country governments will have to deal with trying to co-ordinate, compare and make sense of widely different donor agency approaches, frameworks, indicators, data, terminology, etc. Greater harmonization among donor agencies of their performance measurement and reporting systems, particularly at the country level, would lessen the burden on partner country organizations. However, the extent to which this can be accomplished may be limited given the variation in government-wide performance reporting systems that have evolved in different OECD countries that may dictate the donor agencies' approaches.

12. Conclusions, lessons and next steps

This final section 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. 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.

12.1 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 may be more difficult 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 institution 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-wide programming. 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 partners' efforts or 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 accountability reporting (i.e., holding an agency and its managers accountable for results), and (b) for management improvement (i.e., learning and decision-making) -- may not be entirely compatible with one. 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, or 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-wide 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 measurement and 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 more 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.

12.2 Preliminary lessons learned

Some preliminary lessons learned about establishing effective results based management systems follow:

- *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, analyze and report. 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 RBM 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.

13. 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.

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.