



Key Steps in Designing and Implementing a Monitoring and Evaluation Process for Individual Country Service Agencies

Harry Hatry

This paper¹ identifies key steps in designing and implementing a monitoring and evaluation (M&E) system for ministries and individual government agencies that provide services. These suggestions are intended to apply whether the ministry or agency is in health, education, social welfare, environmental protection, transportation, economic development, public safety, or any other sector. The system might have been ordered or requested by the president or prime minister's office, by a minister, or by any agency head who wants to implement an M&E process. M&E development should focus on providing a process that will yield regular outcome data (in addition to data on the organization's outputs) that can be used by the designing agency and upper-level officials for accountability and, particularly, for managing these organizations, thereby helping officials improve their accountability and services to their citizens.

Prerequisites

There are a number of basic conditions that are vital to successful M&E implementation. First, there is a need for strong, top-level supporting leaders who are willing to provide the time and resources to initiate the M&E process. Second, the presence of at least some staff and enough data collection capacity to undertake basic collection and analysis are critical. Third, government officials must be willing to use the M&E data to help make resource allocation decisions, and not merely use the data to show that the agency is collecting performance data. In addition, there should be enough flexibility for agency managers to make changes where the data indicate that changes are needed. If these conditions are absent, the M&E process may be a waste of time.

Basic Steps

There are a number of basic steps for designing and implementing the process, whether at the ministry or agency level.

- 1. Establish a committee of key agency personnel to guide and oversee the implementation effort.** The committee should include individuals from within the organization as well as representatives from program offices; the budget office; the planning office; the information technology office; the human resources office; and an analytical office, such as a statistical office, if one exists, whose work might be expected to relate to M&E. It is also likely to be helpful to include in the committee an "outside" person who can bring a broader perspective, such as someone from a higher-level budget or planning office.
- 2. Select a small cadre of support personnel** who will be responsible for supporting the implementation of the various administrative and technical elements of the M&E system (under the guidance of the committee).
- 3. Consider meeting in the early stages with members of the legislature** to find out what information would be useful to them for annual budget deliberations or other policy or oversight

purposes. Similar input should be sought from the national budget office.

4. Develop a plan and schedule for implementation keeping in mind it is likely to take three to five years for full implementation. However, some initial useful performance information can probably be expected to become available by the end of the first year.

As part of this step, it is critical to identify who will be needed to undertake each step of implementation. The people selected should be available and committed to providing the time necessary to perform their assignments.

5. Provide training and technical assistance to managers and key professional staff in the basics of M&E. This applies to initial training and for training new people as turnover occurs. The training for managers should provide a thorough understanding of what M&E entails and recommendations for using the information. The training for other staff should focus more on the details on how M&E should be conducted so as to make the information obtained both valid and useful. Multilateral and bilateral funding organizations are potential sources of support for these activities.

6. Ask each program manager within the agency to identify a results-focused mission statement for their programs. These statements will necessarily be general, but they should set the tone and direction for the M&E effort in each program.

7. Ask each program within the agency to select a set of outcomes and outcome indicators that will enable the program to monitor progress in achieving the results identified in its mission statement. Programs could establish their own working groups of key personnel to help select appropriate outcomes and outcome indicators.

The M&E working group might also seek input from representatives of the customers served by the program. For example, the program might sponsor meetings (sometimes called focus groups) with small numbers of its clients to help identify outcomes of importance to them. A public welfare agency, for example, might hold such sessions with its clients; a health agency might hold such sessions with patients in health clinics; and an economic development agency might hold such sessions with representatives from businesses.

The set of outcome indicators should be reviewed at a higher level to ensure that the set contains the important performance information

needed. For example, a ministry should review the indicators selected by each of its agencies. An agency should review the indicators proposed by each of its programs.

8. Inventory existing databases within the agency to identify the output and outcome information already available (or at least readily available) that can be used to monitor progress toward the agency's mission. The sources might be inside or outside the agency, such as data already regularly collected by a central planning office. Many, if not most, countries and their agencies already regularly collect outcome data such as incidence and prevalence of health problems (for example, malaria, tuberculosis, and HIV); road accident fatalities; drinking water and air quality; unemployment rates; school absenteeism and completion rates; and economic development indicators and so forth.

9. For each candidate performance indicator, identify the source of the data, particular data collection procedures, and how often the data will be collected and reported. Some outcomes will require new data collection. Note that an indicator should not be considered final until a feasible data collection procedure has been identified. An agency and its programs should not limit itself to considering only outcome indicators for which it currently has data. Depending on the availability of resources, however, the agency may need to defer beginning new data collection procedures until later years.

In addition to data obtainable through agencies' records, two other major data collection options to consider are customer surveys and trained observer ratings.

Surveys of citizens. Surveys are increasingly being used around the world to obtain feedback on outcomes, including the quality of government services. The feedback can be from samples of households or of customers that have been assisted by the agency (such as patients, students, or businesses). These surveys can provide feedback on a variety of program outcomes, including the surveyed persons housing, employment, or health condition status; ratings of various program quality characteristics, such as the program's timeliness, accessibility, fairness, and ease of obtaining information on the program's services; frequency of use and reasons for nonuse of the service; and extent of having to make inappropriate payments to obtain service.

Citizen surveys can also obtain information on the demographic characteristics of respondents, which helps identify the distributional effects of the program.

Trained observer rating procedures. This option uses personnel to systematically rate physically observable conditions such as the condition of roads, hospitals, health clinics, school buildings, or other public facilities; proximity of households to sources of drinking water (such as standpipes); and the ability of persons with disabilities to undertake basic activities of daily living. Each of these conditions is rated by an observer trained in the use of a well-defined rating scale. The raters visit all, or samples of, the units to be observed, such as the roads, hospitals, or other facilities. The outcome indicator would be the “number and percent of those items rated as being in satisfactory condition.” Ratings obtained at regular intervals enable the program to monitor the changes in each condition.

The procedure has a potentially very useful side benefit to program operations. The ratings of individual facilities or other units that are inspected can be used to identify where work is needed and the extent of that need (for example, is the condition hazardous) so those units with poor ratings (that is, are in poor condition) can be prioritized for correction as soon as possible.

10. Provide for basic analysis of the performance data. At the ministry or agency level, top-level officials should assign an office to be responsible for examining the data collected, identifying unexpected findings, and reporting these findings to the appropriate officials in time to allow the agency to act on them.

The basic analysis for performance monitoring is typically comparisons. Such comparisons include comparing the current reporting period to previous reporting periods; comparing current data to targets; comparing outcomes among different customer groups; and comparing the findings to those in similar programs in other countries. While more sophisticated statistical analysis would be desirable, this first-level, basic analysis should be highly informative and make the monitoring considerably more useful.

11. Emphasize the importance of outcome information, particularly by using the M&E information in decision making. Top-level officials in the agency also need to use the monitoring information collected on outcomes to identify

needed program improvements and training or technical assistance needs, and to draw attention to successful programs. Using M&E data at all levels of the agency, such as planning, budgeting, and program review, highlights the importance of the M&E system.

12. Review the monitoring data and data collection procedures periodically. This review is an appropriate time to introduce improved measurements to ensure the continued reliability and credibility of the performance measurement work. Performance measurement should always be considered to be a work in process.

13. Use the M&E information to lay the groundwork for *positive* incentives for improving the effectiveness of public services. From the beginning, emphasize that achieving good results will be rewarded. Emphasize that a major purpose of M&E information is to contribute to the learning process to help officials and their staff identify what works and what does not.

14. Emphasize the message that outcome information is highly useful for telling what the running outcome score is, but does not provide the background information that tells *why*. The M&E system should not be used as a blame game. M&E information is often perceived by managers as threatening because it has been used to blame managers for poor outcomes without considering why the outcomes occurred. Outcomes are often affected by factors over which managers have limited or no control.

Making the M&E Process Considerably More Useful

Following are six key ways to maximize the usefulness of an M&E process within a service agency:

1. Link key characteristics of the “customers” the agency/program is serving to the outcomes for each customer. Then, report the outcome data for each such category of customers as well as for all customers. This information can be highly useful both for interpreting the data and for identifying what actions are needed.

For example, program managers in the education, employment, public welfare, housing, and health sectors will find it very useful to be able to identify program outcomes by demographic characteristics, such as outcomes by age, gender, income, household size, race/ethnicity, and so forth.

This may suggest the need for program changes for groups that were not sufficiently helped.

For programs for which people are not direct service recipients, such as road maintenance and environmental quality programs, breakouts by key characteristics are still very important. For example, data calculated on number of traffic accidents will be considerably more useful if the outcome data are broken out by characteristics such as cause of the accident, average daily traffic, and location.

A key problem for many countries, particularly poor countries, is that they may not have MIS systems capable of providing such information. Nevertheless, the programs in such countries should do whatever they can to try to obtain and calculate such breakout data, even if the calculations only can be done manually, and even if they can only focus on outcome breakouts for one or two categories.

2. Set targets for the coming year for each output and outcome indicator. A good time to select targets is during the annual budget process. Choosing these targets is more of an art than a science. Setting targets typically starts with the data on the previous years' values for each indicator. These values then can be adjusted for any significant changes expected, such as in funding or personnel. Other considerations are external factors, such as the economy (such as changing export conditions or new businesses that can be expected to increase business revenues or reduce unemployment); technology (that is expected to lead to improvements in service quality); and political changes expected to alter ministry or agency priorities.

Normally the program should propose the targets for its own performance indicators. These targets, however, should be reviewed by upper-level officials to ensure that the targets are neither too easy to meet nor overly high. The danger with targets is that they can become primarily political in nature, making the targets either much too easy to achieve or unrealistic. Targets are more meaningful when based on previous performance and reasonable expectations of the future.

3. Ask agencies and their programs to provide explanations for unexpected results and indicate plans for correcting any problems noted in the explanations. The ministry or agency might set a threshold such as requiring explanations for any performance indicator value that differs by more

than X percent of the expected target value, such as more than 10 percent.

Data on results do *not* tell what the cause of those results is, or what to do about them. Agencies and their programs should be encouraged to seek explanations for unexpected results and develop plans to correct any problems that are identified. A side benefit of such a provision is that managers may welcome the opportunity to formally provide their explanations, fearing that otherwise they will be blamed unfairly for poor results. This may alleviate the "blame" problem noted earlier.

4. Provide for regular, timely performance reporting for review by ministry, agency, and program officials. For budgeting purposes, only annual performance reports may be needed. However, for managerial purposes, this is too infrequent. Generally, reporting should be done at least quarterly. For some indicators, such as those relating to safety and health, reports are likely to be needed more frequently.

5. Make the report format as understandable, readable, attractive, and informative as possible. Too often performance reports are unclear or uninformative or filled with unfamiliar terminology. With today's electronic capabilities, even poor countries have access to a variety of report-generating software.

6. Ministry officials, agency officials, and program managers should be encouraged to hold regular "How Are We Doing?" review sessions soon after the latest performance report has been issued. The ministry or agency head meets separately with each major agency program or division to discuss the performance data, identify where the outcomes are going well and where not well, and suggest ways to correct poor past results. In subsequent meetings, previous decisions should be followed up on to determine the extent of progress that has been made in correcting previously identified problems.

Incorporating Program Evaluations

The monitoring process described above provides regular measurement of the performance of programs. Such information does not tell *why* the program has been successful or unsuccessful. The explanatory information suggested above provides first order information on the "whys" and what

might be done to improve performance. However, for programs that involve large expenditures and/or are of major importance, more in-depth program evaluation is desirable. In general, program evaluations tend to be expensive and may require long periods of time, perhaps years before the findings are available.

Each year, it is important for the ministry and each agency to establish a schedule of which programs will receive an in-depth evaluation. The choice should depend on criteria such as major uncertainties and lack of information regarding program success; cost of evaluation; time required

to complete the evaluation and therefore the timeliness of the findings; and likelihood that the evaluation will provide useful information for later decisions. For some programs it may not be feasible to obtain the needed information. An initial step is to assess whether these programs are good candidates for evaluation.

It is also critical to select who (what type of organization) will conduct each evaluation. The organization conducting the evaluation should be reasonably independent, if for no other reason than to achieve external credibility for the findings. Few agencies or ministries governments

Box. 1: United States National Park Service

The United States National Park Service (NPS) is responsible for 392 park facilities throughout the United States. NPS, as well as all Federal programs, has been required to provide Congress with annual performance information. The United States Office of Management and Budget left it to each Federal program to implement its own performance measurement process. NPS first identified its mission: To preserve the natural and cultural (historical) resources for posterity and to provide those resources for the enjoyment of U.S. citizens.

NPS decided to develop a performance measurement process that would: (1) help identify budget needs and provide justification for funding both individual parks and the park system as a whole; (2) help the individual park superintendents manage; (3) provide fair comparisons across parks.

NPS identified six basic outcome areas to be measured on at least an annual basis: (1) condition of natural resources; (2) condition of cultural/historical sites; (3) park-user satisfaction; (4) condition of noncultural facilities; (5) safety of park users and employees; and (6) financial management, including park efficiency.

For each outcome area, NPS identified outcome indicators. Measurement of the indicator values has presented some difficult measurement questions, especially given the diversity of the 392 facilities. NPS developed a “scorecard” containing about five key outcome indicators for each of the six outcome areas.

For park-user satisfaction, NPS contracted with a university to develop a process, administered by each park, to annually survey random samples of park users and ask them to assess a number of key park characteristics. Most of the other outcome data for the other five outcome areas are provided by each park.

The scorecard process has evolved over a number of years and is still evolving as measurement procedures have improved. Each park superintendent is annually given its own scorecard results, along with a comparison to the average for all parks and the average of the other parks in its region. Each park can also compare itself to any other parks it believes are similar.

Thus far, the annual process appears to have been more useful for accountability and budgeting than for park management, partly because of the lag time before the scorecard data become available to the parks. However, much of the data can be used by individual parks to identify actions needed throughout the year. For example, the data on safety (accidents, injuries, crimes), feedback from park users, and facility condition data (based on regular inspections) are routinely considered and used for allocating resources. The parks are encouraged to use the scorecard information as part of their justification for funding requests as part of the annual NPS budget process.

The outcome indicators and process have been evolving over the years and will likely continue to be improved as new measurement procedures become available.

Source: Compiled by author.

are likely to have the in-house capability for such program evaluations. Almost certainly the evaluations will need to be conducted by outside organizations such as universities or specialized private nonprofit or for-profit organizations. Because of their costs, the ministry or agency might seek funding from one of the international multilateral or bilateral funding organizations, particularly for evaluating programs for which the funding organization has provided support.

An important additional step is to ensure that the completed evaluations are professionally reviewed to assess their quality. Most program evaluation reports should be made available to the public. This step itself can help encourage the program evaluators to ensure their work is of reasonable quality.

Uses of M&E Information

M&E information has many uses, including identifying shortfalls in outcomes, thereby enabling service improvement. M&E data can also help to formulate ministry, agency, and program budgets and help to justify budget requests. Program outcome information, whether obtained through regular performance monitoring or from program evaluations, should be linked to program cost information as part of the budget process. The outcomes expected to be achieved by the budget requested should be a major factor in budget decisions.

The information produced by M&E systems is also useful for helping to allocate and prioritize resources throughout the year, such as adjusting the assignment of work or staff to different locations or customer groups based on performance data. For example, outcome indicator data can identify: areas most in need of road repairs; emerging health issues; types of housing shortages; traffic accident locations; and locations with the most water or air quality problems.

M&E systems can also develop incentives for agencies and programs based at least in part on success in achieving outcomes. Monetary incentives can be expensive and are not likely to be appropriate for many countries—at least not until considerable experience in M&E has been gained. However, nonmonetary incentives can be considered once the M&E process is in place. For example, as noted earlier, recognition awards

could be given to agencies or programs that have met (or exceeded) their targets.

M&E information provides the basis for developing multiyear strategic plans for ministries and agencies. The latest values for key outcome indicators can be used to establish the baselines and subsequent out-year targets. In later years, the annual data from the M&E process on these key outcome indicators can be compared against the targets in the multiyear strategic plan. Such information indicates whether actions are needed to stay on the plan, whether plan revisions are needed, or whether the plan is no longer feasible.

The information produced also helps to communicate with the legislature and citizens. Performance reports that provide data on the ministry's or agency's major outcome indicators

Box 2: Mexico's Ministry for Social Development

Mexico's Ministry for Social Development (SEDESOL) found that the development of a results-based management system takes time and is a work in progress. In 2007, after earlier piloting a results-based monitoring and evaluation process in four programs, SEDESOL began implementation of a more comprehensive process. The process combined results-based monitoring, evaluation, and budgeting. The ministry early on established an office with extensive evaluation capacity, providing a basis for the later results-based process, which included a combination of monthly, quarterly, and annual data reporting.

As of 2009, SEDESOL had 500 indicators of which 112 were linked to results. One hundred seventy of the indicators measured the quantity, efficiency, and quality of the services delivered by programs. The remaining 218 indicators measured the activities used to produce these products. However, 40 of these indicators were selected as priority indicators for use in the 2010 budget preparation and subsequent monitoring. These 40 were selected by the Ministry of Finance with input from the programs. The Office of the President also monitors the indicators and goals of sectoral programs such as SEDESOL.

Source: Mexico Ministry of Finance, and World Bank. 2010. "Mexico Quality of Public Expenditure." Note No. 11, June.

are likely to be of considerable interest to legislators and citizens. Such transparency can be helpful in obtaining citizen support for the ministry's and agency's work.

Ministries and their agencies may, understandably, fear that showing bad news (such as worsening outcomes) will only bring them grief from the legislature and public. However, showing only outcomes that look good is likely to lead to a lack of credibility. Problems can be alleviated if the ministry or agency, through their programs, provides explanations for poor outcomes and indicates what it plans to do to correct the problem.

Intergovernmental Concerns

Depending on the level of decentralization in the country, a considerable amount of the data may come from local or regional/provincial government bodies. For example, to the extent that education and health services are decentralized, these lower levels of government should be monitoring the outcomes of these services.

The ministry or agency will likely need to help lower-level governments develop their capacity to operate their own M&E systems. The central ministry or agency might need to provide guidance, guidelines, and financial support. The national agency will need to check the quality of the performance data from lower-level governments; this means providing training and technical assistance, but also periodically auditing at least a sample of the data and data collection procedures used by lower levels of government.

Cost of M&E

The added cost will depend considerably on the extent to which each agency program is already collecting reasonably reliable performance data and if the ministry, agency, and program already have personnel that can assist in implementation. Many agency programs are likely to be already tracking a number of outcomes, which can be used as a starting point for the M&E system.

The largest added costs will likely include the costs for any new personnel needed (such as for analysis); training, including start-up, continuing, and replacement training; and for any new data collection procedures, such as the cost of surveys. Household surveys, in particular, will probably need to be administered by an outside business,

university, or the national statistical office. Added costs could also include costs for a new computer system and technology to process the substantial amounts of data, as well as in-depth program evaluation costs. In-depth program evaluation is usually conducted by an outside organization such as a university or specialized private firm.

The position of many governments in developed countries that are introducing M&E systems has been that monitoring is a basic management function. The cost of ongoing implementation should, therefore, be primarily covered by the ministry, agency, and the program's own budget. Initial start-up costs are likely to be more of a problem for many ministries and agencies, however, and for these costs funding could be sought from the central government or from donor organizations.

Final Thoughts

Monitoring and evaluating the results of public services is both common sense and good management. It needs to be recognized that performance data obtained from M&E systems do not replace the need for judgments by public officials. M&E information can only serve as one input, albeit a major input, for decision making. Many other factors will also need to be considered.

Implementing M&E successfully and usefully in a ministry or agency requires commitment and leadership from high-level officials in these organizations. And it takes time and special resources. If the organization's climate will not likely permit meaningful use of M&E information, the effort will be a waste of money. However, if the climate is favorable, the ultimate gains should be considerable in improving the organization's services to its citizens.

Further Reading

This brief note provides an overall picture of steps for implementing an M&E process in a ministry or agency. Considerably more detail is provided in the following literature.

Gorgens, Marelize, and Jody Zall Kusek. 2009. *Making Monitoring and Evaluation Systems Work*. World Bank.

Hatry, Harry P. 2006. *Performance Measurement: Getting Results*, Second Edition. Urban Institute.

Mackay, Keith. 2007. *How to Build M&E Systems to Support Better Government*. World Bank.

Zall Kusek, Jody, and Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. World Bank.

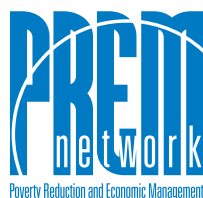
About the Author

Harry P. Hatry is a *Distinguished Fellow* and *Director of the Public Management Program* for the *Urban Institute* in Washington, DC. He has been a leader in developing performance management and performance management procedures for federal, state, and local public and private agencies. His book, *Performance Measurement: Getting Results, Second Edition*, is widely used and has been translated into two other languages. He has provided assistance on M&E to the U.S. Departments of Education and Justice, Health and Human Services and the

Environmental Protection Agency, and has participated in projects in Albania, Colombia, Hungary, Indonesia, Pakistan, and Thailand.

Note

1. The author thanks Markus Goldstein (Senior Economist, Africa PREM), and these members of the Poverty Reduction and Equity Group: Helena Hwang (Consultant), Philipp Krause (Consultant), Gladys Lopez-Acevedo (Senior Economist), Keith Mackay (Consultant) and Jaime Saavedra (Acting Sector Director) for their comments. The views expressed in this note are those of the author. To access other notes in this series, visit www.worldbank.org/poverty/nutsandbolts.



This notes series is intended to summarize good practices and key policy findings on PREM-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. PREMnotes are widely distributed to Bank staff and are also available on the PREM Web site (<http://www.worldbank.org/prem>). If you are interested in writing a PREMnote, email your idea to Madjiguene Seck at mseck@worldbank.org. For additional copies of this PREMnote please contact the PREM Advisory Service at x87736.

This series is for both external and internal dissemination