Is Evaluation Embraced in Turbulent Times of Economic Crisis?[1]

LINDA G. MORRA IMAS

If there was ever a time when decision makers would gladly embrace evaluation, it would seem to be in the turbulent aftermath of the Global Economic Crisis of 2008. Even as we stand well into 2010, the effects stretch on. It is clear in hindsight that the data to alert decision makers of the coming crisis and to take action were available, but not used. While many experts misjudged the timing, speed, and severity of the crisis, the economists who in fact did predict it, did not have their voices heard. As put by the World Bank's Chief Economist, Justin Yifu Lin (2009): "Though many knew that the risks of a severe financial crisis were mounting, the necessary changes in policies and practices, both in mature economies' financial sectors as well as in many new emerging markets, were stymied by procrastination during the 2003–2007 boom. No one was willing or capable of taking the punch bowl from the party and the global institutional set-up did not have the leverage to do so" (p. 29). This chapter begins by examining the issue of whether the turbulent aftermaths of economic crises increase the demand for evaluation from policy makers. It then moves beyond this issue to the premise laid out by the editors of this book, that in turbulent times "old solutions, old instruments, and old institutions" are not sufficient…and "marginal changes will not work". Applying this to the aftermath of the economic crisis of 2008, we ask whether it is less a question of how to use evaluation to improve ongoing programs and policies, and more an issue in times of economic crises of how evaluation can help a needed paradigm shift and how a paradigm shift also may influence evaluation.

This chapter develops five points—

1. turbulent economic times can be an impetus for evaluation, but there are other factors that can come into play, especially in more stable times;
2. often turbulent times increase demand for information on performance of government-funded programs, but the demand is for results-based performance monitoring systems that can be linked to budgets;
3. effective response to economic crises requires sound data from *integrated* systems of monitoring and evaluation (M&E);

[1] This chapter is based on a keynote speech, "Embracing Evaluation in Tough Economic Times", given by the author to the Ontario Chapter of the Canadian Evaluation Society in Toronto, October 1, 2009. It is to be published as a chapter in a forthcoming book edited by Furubo, Jan-Eric, Ray C. Rist, and Sandra Speer. Expected publication date is 2011. *Evaluation in Stable and Turbulent Times*. Transaction Publishers. This version is intended for the limited use of IPDET 2010 participants and should not be copied or further distributed.

4. integrated M&E systems must take into account incentives *to use* the data, but a decade of lessons learned on promoting evaluation use points to *need for new strategies* even in stable times; and
5. especially in times of turbulence, monitoring and evaluation specialists must demonstrate how they can bring information to bear on real-time issues facing decision makers.

Turbulent times and demand for evaluation

Economic crises initially may appear to increase demand for evaluation. In a global financial meltdown, the first reactions, evaluatively speaking, often are kneejerk ones-- How did we get into this mess? Where did the money go? Who or what is to blame? How could the economic crisis have been avoided? These evaluative post-mortems are a temporary flurry of largely unproductive activity.

Fortunately, the move to the second stage is fairly rapid. During this stage, the questions are: Given that we are in this crisis, what can we do about it? How can we recover quickly and get back on our feet? What are the lessons learned from the other times something like this has happened? We may see a more productive flurry of evaluative activity within organizations to describe what was done in prior crises, determine how well or poorly it worked, and identify the broad lessons learned that are applicable to present circumstances. The Independent Evaluation Group of the World Bank Group, for example, issued in December 2008 an analysis of lessons from World Bank Group responses to past financial crises. Similar reviews are undertaken for many organizations.

But is there more sustained interest in and promotion of evaluation in turbulent economic periods? In their Introduction to this book, Furubo and Speer suggest that decision makers may turn to the known subject experts for advice on best alternatives in these times, and not to evaluators. Still, there is evidence that for some, there is more interest in evaluation. Certainly this has been the case for the U.S. Government Accountability Office, the GAO. The American Recovery and Investment Act, passed by Congress and signed into law in response to the crisis, might well be considered a full employment act for evaluators—at least for GAO evaluators. It has GAO conducting bimonthly reviews of selected states' and localities' uses of stimulus funds; reviewing and commenting on reports on the number of jobs created or preserved; doing special studies on areas such as trade adjustment assistance and efforts to increase small business lending; and monitoring long-term effects of the economic downturn on states, such as health care costs. It is interesting to note the types of evaluative information requested. Perhaps key among them and containing major political challenges are how the numbers of jobs created and jobs preserved are defined and measured.

It seems common sense that evaluation would be especially important in times of sustained economic crisis for national governments. When resources are scarce and competition for them increases, heightened interest seems reasonable in knowing which programs are achieving intended results with a high cost-benefit ratio and which are not. Such information could be perceived as helping officials make better decisions about which programs to continue or discontinue or to give funding increases or decreases.

The case of the U.S. 1980 recession

Putting aside "coulds and shoulds", has this been the general case? We can turn back the clock and go to 1980 in the U.S. where the 1979 energy crisis was followed by a 1980 recession, creating a long period of turbulence. The economy was suffering from slow growth, high inflation, rising unemployment, and unprecedented interest rates. The states were in crisis and experiencing serious budget shortfalls and slashing budgets. In this tough economic climate, Ronald Reagan was elected to his first presidential term and took office in 1981. In taking office, Reagan referred to the economic crisis he inherited as "the worst economic mess since the Great Depression." His election was a call for paradigm shifts in federal policies. Reagan ran on a pledge to reduce federal taxes, reduce federal regulation, and reduce the federal workforce. He aimed to reduce the federal role in government and increase that of the states through the use of block grants. The Deficit Reduction Act passed in his administration looked for targets for savings in government operations. So the question is: did evaluation flourish in this period of economic turmoil?

Fortunately, there are data to examine. The GAO's Program Evaluation and Methodology Division (1987) looked at the extent to which federal non-defense evaluation program activities changed between 1980 and 1984. The GAO found substantial overall *loss* in (1) the number of units engaged in program evaluation; (2) in fiscal resources, professional staff and products; and (3) in information about the extent and nature of program evaluations themselves. Funds for evaluation units decreased in constant dollars 37 percent, but the departments and agencies in which these evaluation units were located received an average 4 percent increase over the same period.

Closer inspection of the nature and scope of evaluation activity in 1984 compared to 1980 concerned the GAO even more in that there was a implicit policy shift away from full-fledged evaluations to internal studies—low cost, short turn-around, and non technical—done for program management. The GAO concluded that these kinds of rapid evaluations cannot typically present strong information on program results and pointed to a reduced future availability of adequate information for congressional oversight. All this makes the point that turbulent economic times may result in paradigm shifts for evaluation as policies shift as a result of the pressing need for alternatives. But these shifts are not necessarily a boon for evaluation.

The cases of Australia, Great Britain, and New Zealand in the early 1980s

While interesting, however, the U.S. in the Reagan era is an "N of 1". Staying in the early 1980s, a look at the case of Australia yields a picture similar in that turbulent times led to a paradigm shift, but in this case, it was a shift to evaluation, rather than away from it. Much has been written about Australia's warm embrace of evaluation in 1983 when a new government faced a difficult macroeconomic situation. According to Mackay (2007), evaluations were viewed as providing the necessary in-depth, reliable information on the efficiency and effectiveness of government programs. In a policy shift, a new requirement called for formal portfolio evaluation plans to be submitted annually to the minister for finance. Furthermore, every program was to be evaluated every 3-5 years. Reportedly, significant use was made of evaluation findings to support the cabinet's budget decision making. Performance information, however, was left to line departments.

However, Australia does not appear to be the norm. More frequently in the face of sustained economic crises, rather than embrace evaluation, governments appear to shift to emphasizing results-based performance monitoring systems. Great Britain is an example. In the early 1980s, Britain also had a tight fiscal policy and concern with administrative costs. Gray and Jenkins (1989) reflect on the British experience in relation to policy evaluation in this time of crisis. Rather than a growth in evaluation in response to this crisis, the reaction was that Britain moved away from policy evaluation and to performance measures and performance indicators. Gray and Jenkins argue that the way governments develop and use evaluation is conditioned not only by the macroeconomic climate in which they operate, but also by the ways they choose to respond to it, and by the microeconomics of the budgetary process. In Britain the response was to move from "top-down rationale systems of evaluation to more pragmatic, bottom-up concerns with performance measures and performance indicators". This was a paradigm shift for evaluation in Great Britain.

The current global economic crisis does not appear to be resulting in increased use in Great Britain of policy or program evaluations, although a change of performance indicators from shorter term to longer term was recently called for by former Prime Minister Gordon Brown. Ahead of a meeting of EU leaders, Brown told reporters in Brussels: "I think for the future it is really important now that we have proper systems of bonus and reward that reflect not short term deals but reflect long term success". The new message was one of tying performance indicators to personnel management compensation to create the right incentives for desired performance.

In another example, New Zealand, the economy was in dire condition in the early 1980s. Lack of action was not an option. New Zealand chose to design and implement a performance management system and procedures. Its system at that time was characterized as heavily output-based with weak links to evaluation. So again, turbulent times appear to have been an impetus.

A decade of crises 1993-2003 and beyond

The Independent Evaluation Group of the World Bank noted that in the decade of 1993-2003 there were crises in 18 countries, in some of them, more than one. In looking across these countries the only constant is the lack of evidence of clear link between the economic or financial crises and embrace of evaluation across these countries. For example, the Asian economic crisis is credited with facilitating evaluation practices in Korea's executive branch. In Malaysia, however, its efforts in 1990 to build the nation and increase its global competitiveness, as laid out in its Vision 2020, are associated with its introduction of a performance monitoring system linked to budgets. In some countries with economic crises, there was no embrace of any form of performance monitoring or evaluation.

In Canada, a 1994-95 program review that is credited with solving a budget crisis, perhaps along with budget surpluses that were the rule from 1997-1998 onward, can be considered the impetus for the 2000 introduction of the Results Management Accountability Framework and subsequent Management, Resources, and Results Structure Policy of 2005. Government focus on good planning, performance assessment, ongoing expenditure review, and Parliamentary pressure for transparency on the use of public funds and for efficiency and effectiveness seem the continued system drivers.

In sum, it is clear that factors other than turbulent economic crises can lead to national governments walking down the aisle hand in hand with performance monitoring or evaluation. Sometimes it is a government committed to a vision of improved public administration and transparency; other times it is donor support for statistical or evaluative capacity building, or even formal requirements for accession to the European Union. And sometimes it is evaluation that is embraced but it seems that more often, it is a form of performance monitoring.

Emphasizing the “M” in “M&E” in turbulent times

When an economic crisis does serve as an impetus for a policy shift for evaluation, it seems that it is the “M” that is emphasized, rather than the “E’, and often it is linked to the budget process. It may be called results-based monitoring, and if it links with a budget, then it is referred to as results-based budgeting, to emphasize the focus on results, rather than outputs. Sometimes it is referred to as results-based management. But by any of these names, it is often the result of a siren call to governments faced with strong sustained pressures not only to account for the billions they spent by showing what the money bought, but also to identify the results they have obtained with the expenditures. It can help government finance units deal with the competing demands on the budget.

In his inaugural address, during the height of the turbulent economic crisis, U.S. President Obama thrilled those in the evaluation profession with his message for a policy shift to emphasize program performance and link it to budget actions. His words were: “The question we ask today is not whether our government is too big or too small, but whether it works -- whether it helps families find jobs at a decent wage, care they can afford, a retirement that is dignified. Where the answer is yes, we intend to move forward. Where the answer is no, programs will end. And those of us who manage the public's dollars will be held to account -- to spend wisely, reform bad habits, and do our business in the light of day -- because only then can we restore the vital trust between a people and their government.” This policy shift to RBM may be welcome and timely, but is it a paradigm shift that is likely to lead to breakthroughs and alternative directions for government to guide the future? Or to the contrary, does it support Furubo and Speer’s hypothesis that in turbulent times, evaluation will likely have a less dominant and more indirect role in contributing to new knowledge structures?

Issues in using results-based monitoring systems

There is a growing literature on issues in the use of results-based performance monitoring and performance budgeting systems. At the heart of the matter is the difficulty making decisions from limited performance data that are essentially quantitative program scorecards. Such scorecards hold out the promise of making funding decisions easy, but they do not always deliver. Monitoring data give ongoing information on the direction of change and the pace of change and the magnitude or extent of change. Such data can identify outliers, those performing better than expected and those performing worse than expected. Monitoring data also let us know what is typical. But these types of data can raise more questions than they answer.

To illustrate, we may know how a program fared on its key indicators in meeting its targets but we do not know why this is the case. Was the failing program under-resourced in the first place, poorly implemented, ill-conceived in its theory of change, lacking relevant partners, or a victim of over-ambitious objectives? Or was it simply not a cost-effective intervention given the objectives? Could the successful program have been more successful? Were there significant natural variations in implementation that are associated with outcome gains? Did the program itself make a difference in outcomes, independent of other influences? What elements of the program appeared to have the greatest influence on achieving the outcomes? Did the program's long-term impacts justify the costs? Should the pilot be expanded? To what extent was underfunding a likely cause of poor program results? Were there more effective and/or less expensive ways to obtain the same result? Performance monitoring systems cannot answer these questions for budget examiners. Evaluation is needed to explain the "whys" and determine attribution.

A concern is that results-based monitoring systems tend to focus on a limited set of shorter-term outcome indicators, sometimes of dubious validity, and therefore they may lead to erroneous conclusions about program success or lack thereof. Outcome indicator data may only reflect weak correlations but are too often interpreted as if they were the direct and exclusive effects of program interventions. We know that evaluation is needed for attribution of causality for observed changes. But perhaps most importantly, we need evaluation to answer the "why" questions that monitoring data raise—to provide the explanations for what is observed.

A case illustration

So what can happen when limited performance monitoring information is linked to the budget? The U.S. again serves as an example. Starting with the 2004 budget, under the Bush Administration, the U.S. Office of Management and Budget (OMB) began using the Performance Assessment Rating Tool, known as PART, to help allocate budget resources. Like other governments, the U.S. endorsed performance measurement linked to performance budgeting because of the promise held for determining which government programs produce results and thus should receive budget increases. PART was intended to complement the Government Results and Performance Act of 1993 (GPRA) in which agencies develop 5-year strategic plans (updated every 3 years), annual agency performance plans, and annual agency performance reports for the prior fiscal year. PART yields ratings for each program's (a) purpose and design, (b) strategic planning, (c) program management, and (d) results. Each federal agency provides material to OMB to answer each item on a rating questionnaire. OMB budget examiners then score each item and summarize the ratings into an overall "effectiveness" rating. PART findings and performance information is intended to justify funding requests, management actions, and legislative proposals. OMB has reviewed about 20 percent of the programs each year.

Issues surrounding PART have included the lack of meaningfulness of the summary "effectiveness" measure used and need to increase focus on the "outcome" or results measure, disconnects between agency strategic priorities (GPRA) and the program outcomes defined by PART, competing demands on the time of the OMB Examiners, inconsistencies in how performance measures (outputs and outcomes) are defined for programs, lack of outcome specification, over-simplified responses, large number of programs receiving ratings of "results not demonstrated", lack of funding to support agency data collection, time needed for outcomes to be realized, and need for disaggregated outcome data by geographic and demographic characteristics and by federal government organizational unit (e.g. region, facility, office).

The largest issue, however, has been use. Executive branch management, funding, and authorization decisions have not been regularly based on PART. And in the Congress references to PART have been rare; PART has not been the basis for legislative action, even in more stable economic periods when there is more acceptance of incremental change.

This situation is not unique to PART. Gilmour and Lewis (2006) conclude from their review of the literature that there is little systematic evidence thus far that performance budgeting, as implemented in U.S. states and cities has had a major impact on budgeting decisions. Richards and Goh point to a recent survey in Canada of 117 deputy ministers and chief administrative officers that found that only 17 percent used performance measures to make key decisions. Clearly, within a public sector environment, politics, strategy, and organizational culture come into play.

A system, however, that is not used will not be long sustained. The literature on incentives and disincentives to sustaining performance systems in general is large. There are diagnostic instruments to identify the strengths and weakness of a performance system, matrices of possible actions to strengthen such systems and lists of critical components of sustaining use of these systems.

A strong message of this chapter is that embracing results-based monitoring systems or results-based budgeting systems may be passionate actions for organizations, especially in turbulent times of economic crisis. But they may be short-lived flames unless a way is found to simultaneously embrace evaluation for a lasting relationship in good times and in bad times.

Results-based monitoring is a global phenomenon, not just in developed countries, but in countries as diverse as Chile, Colombia, Mexico, Sri Lanka, South Africa, and Uganda. There is much value in such systems as managers need to obtain quick and continuous feedback on how programs are progressing in implementation and how they are performing in reaching their targets. Results-based monitoring systems are also extremely useful to evaluators as they design evaluations and look at what monitoring data exist that can be used, especially baseline data. In fact, to return to our discussion of Australia, its system did not ultimately survive. Reasons for its decline generally include the burden it imposed and the uneven quality of the evaluations conducted by line departments.

Evaluations can take time and often are expensive. They have little purpose if not used. And if done as a matter of routine on all programs, they may swamp the available demand. More work needs to be done to determine when it is appropriate to settle for indicators of achievement, when a rapid review will suffice, and when a rigorous evaluation is needed. But a key reason given for the demise of Australia's system was the lack of performance information to facilitate the planning and conduct of evaluations. Many "independent" evaluation units only conduct ex poste evaluations in the belief that their independence would be compromised if they offered any guidance to operations staff in the policy, program, or project design phase when the design for M&E should also be built. Yet, ex poste evaluation is dependent on the adequacy and quality of the data identified during the early life of the intervention as important to collect. Ex poste evaluations are especially problematic when the intervention lacks baseline data or the indicators selected for tracking outcomes are not reliable or valid measures. While there is justification for independent evaluation units, evaluation needs to take place in all phases of programs, but especially as indicated by Rist (2006) when (1) regular measurements of key indicators suggest a sharp divergence between planned performance and actual performance, (2) performance indicators consistently suggest weak or no results from an initiative, or (3) similar projects, programs, or policies are reporting divergent evidence of outcomes.

Integrating Results-Based Monitoring and Evaluation Systems

We ask next where we can find the examples of national monitoring and evaluation systems working well together in integrated fashion. Seldom is the same unit responsible for both functions. And seldom is monitoring accorded the same status as evaluation. How many colleagues do you know that proudly call themselves "program monitors"? One is a management function coming out of quality management and dedicated to learning. The other is a discipline coming out of social science research and auditing that has accountability foremost.

While not the norm, Chile is an example of countries with integrated monitoring and evaluation systems. Chile's system is a highlight in this regard. It has six main components. The first is an ex ante cost-benefit analysis of all investment projects by the planning ministry. All other components are based in the Ministry of Finance (MOF). The second is performance indicators and the third, comprehensive management reports prepared annually by ministries and agencies which include their objectives, spending, and performance. Fourth are evaluations that follow a standard format, including logframe analysis, a desk review and analysis of existing data. Mackay (2007) indicates these comprise "rapid reviews". Rigorous impact evaluations, introduced in 2001, are the fifth component, and sixth are comprehensive spending reviews which analyze all programs within a particular functional area. Indications of poor program performance from performance data are used as a trigger for an investigation of the causes. Either a rapid evaluation or impact evaluation is conducted. MOF identifies the programs to be evaluated and is highly selective in doing so, anticipating the information needs that will arise in the budget process. High use of M&E information in the budget has been reported, but it is also acknowledged that it is one input among others used in decision making. The M&E information also is viewed as successfully driving management improvements in sector ministries and agencies.

In Chile, standardized terms of reference are used for the evaluations which are contracted out. They are conducted within tight time constraints to ensure they can feed into MOF's budget analysis and decision making. Their cost, and indeed the cost of the full M&E system is low; perhaps, some suggest, too low, as sophisticated impact evaluations can be quite costly.

Pfeiffer (2007) indicated at a World Bank session given on PART that OMB had wanted to promote evaluation to measure and improve program design, implementation and effectiveness, including cost-effectiveness. But the evaluations are not used enough to assess impact and improve performance because "decision makers do not appreciate and consequently do not routinely invest in evaluations" and "technical complexity can make them hard to understand and thus undermine confidence in results." This is not a new refrain when it comes to evaluation use, but it is a frustrating one. Add to it the complaint that they take too long.

We know much about increasing use of evaluation findings at the individual program or project level-- from building early ownership of the evaluation to message-oriented report writing to early planning of dissemination activities. The literature on increasing use of evaluation findings is a vast one and the main points well-known from building early ownership of the evaluation to report writing (see, for example, Chelimsky (1987), Weiss (1999), Patton (2008), and Morra Imas and Rist (2009)). Gray and Jenkins (1989) believe that the debate on the use of evaluation often starts from the normative assumption that it is a rational and rewarding activity for governments to pursue. Failure to use evaluation is assumed to indicate political organizational inefficiencies that need to be corrected. But the authors indicate this is too simplistic a view. Picciotto (2005) believes that "accountability" is critical for use: "If evaluation is only about learning, it does not make authority responsible. If it must churn out lessons to justify itself, it will generate an oversupply of simplistic and pious exhortations and platitudes. Worse: evaluators that do not encourage accountability for results fail to provide incentives for learning."

But the literature on *integrating* the "M" and the "E" of M&E, and achieving good use of both is comparatively thin. In their book, *Ten Steps to a Results-Based Monitoring and Evaluation System*, Kusek and Rist (2004) notably do address the issue. But only one of the steps is devoted to the "E" in M&E—some 16 of 170 pages of text. Looking at the Morra Imas and Rist (2009) textbook, it is the same treatment, but reversed.

Strategies for Integrating the "M" and the "E"

So what might be done to integrate M&E? Eight paradigm-shifting strategies for M&E emerge.

1. *A Seat at the Table for M&E*. Use of M&E in the public sector must go beyond the steps discussed to become institutionalized and make a difference. It is everyone's job in general but no one's specifically to bring the lessons of evaluation to the table and ensure that they are at least considered in policy making. This situation needs to change and accountability must be assigned. It is encouraging that President Obama has followed up on his inaugural address by appointing a Chief Performance Officer who also serves as Deputy Director of OMB. Still, it is not a Cabinet level position, nor, as its title would indicate, is it clearly encompassing of evaluation. It is also promising that Canada has moved to requiring evaluation chiefs in every ministry. But the paradigm shift has come from South Africa. This country has recently appointed a Minister of Performance Monitoring and Evaluation. It is the first country in the world to have an M&E czar with a seat in the Cabinet. This means that there should be routine consideration of the knowledge M&E can bring to bear on a specific issue and routine consideration of need to specify M&E provisions and earmark resources when new policies and legislation are under serious consideration. Realistically, such appointments may or may not be successful depending on the extent of streams of evaluative knowledge that have been developed and the knowledge of these streams that the individual comes with or quickly can be provided. If knowledge is not there, there may be only weak thrust and weak effect. But at the same time, one cannot let the imperfect be the enemy of the good. To institutionalize routine consideration of M&E data, countries may need not only good results-based monitoring systems that are well-linked to evaluation, but also a Minister or Secretary responsible for M&E with a seat at the decision making table.
2. *Monitoring and Evaluation Policy Task Forces*. Another way to raise the visibility of available M&E information and promote routine consideration of M&E data is the establishing of Monitoring and Evaluation Policy Task Forces by strong national evaluation associations. One relevant example is the Evaluation Policy Task Force established by the American Evaluation Association. As described by Chelimsky in King et.al. (2009), this is an effort "about helping people use evaluation to answer important program and policy questions in governments; about making it practically and ethically feasible for evaluators to do their job; and about getting a seat for evaluation at tables where decisions are made about evaluation policy in government" (p. 243). As Chelimsky notes, particularly in turbulent times: "We need to convince people—and now seems a very good time, with government accountability in ruins everywhere we look—that evaluation should be built into major programs (both old and new), so that when we want to understand what has or hasn't been achieved, or whether a program's participation rates, say, warrant program continuation, or whether what works in one place will also work somewhere else, we won't have to start de novo, without baseline data against which to compare"(p. 243). Ideally, such efforts would be titled "*monitoring and evaluation policy task forces*" to emphasize the integration of the two, but the premise is that many other national or regional associations could form M&E Policy Task Forces that work to be heard at the policy-making table.

3. *Focusing training programs on both the "M" and the "E"*. Another strategy concerns evaluation capacity building programs and university training programs. These programs also need to focus on M&E, not just the "M" or the "E". Review of the curricula of ten master and/or doctoral degree programs in the U.S. in public policy, public administration, and evaluation, finds none that clearly address the "M" end of the spectrum. Indeed an expanded search found only one university that focuses specifically on monitoring in addition to evaluation. This exception is Jimma University in South Africa which offers a Master's in Health Monitoring and Evaluation. It features such courses as the Theory and Practice of Monitoring and Evaluation. One might wish that more universities offered comprehensive training for evaluators. But they continue to operate in the main as if no one needs serious training in developing, implementing, and using results-based monitoring systems. It seems as though there is general expectation that these skills will get "picked-up" along the way, perhaps with a few hours workshop to explain terminology. And then we expect integrated M&E systems to just happen. We should not be surprised when it does not work this way.
4. *Good practice standards for M&E.* Standards, put most simply, are a set of rules in a published document that are to be followed consistently in order to ensure quality. For example, the OECD Development Assistance Committee (2010) has recently issued quality standards for development evaluation that are intended to serve as an incentive and inspiration to improve evaluation practice. While there will never be only one way to build an effective M&E system, guidance in the form of standards for good practice M&E systems would be a good contribution for all. This effort should include guidance on selecting what level of review and evaluation is needed for a given program. Those involved in monitoring and evaluation will be following the Canadian Treasury Board's approach to performance management with great interest. It is an experiment in a risk-based approach to evaluation with high-impact and high-cost programs subject to more rigorous evaluation. Recognizing that rigorous evaluations take time, it promises to avoid placing a high burden on most minor initiatives by using value-for-money reviews. Its review of government spending and results organized by outcome, rather than department or program perspectives, is especially promising. This is a paradigm shift for evaluation.
5. *Competencies for M&E*. The movement to develop competencies for evaluators has gathered steam due to a desire to increase professionalism and improve the quality of evaluation products. Competencies are sets of behaviors (knowledge, skills, and attitudes) that describe excellent performance in a particular work context. They are often linked to a set of standards. Morra Imas (2010) has described the efforts of the International Development Evaluation Association (IDEAS), a global association for individuals working in evaluation in the development context, to develop competencies for those who conduct or manage development evaluations. Many regional and national evaluation societies, bi-lateral development organizations, universities, and non-profits have established or are drafting evaluation competencies. The same points that have been made previously on integrating the "M" and the "E" hold for organizations in developing competencies for those who work in evaluation. If we expect complementarity and hope for staff fungibility, then competency programs need to identify the knowledge, skills, and attitude sets for both.

6. *Expand the types of groups who need training related to M&E to the demand-side.* Capacity building statements and positions, for example the recent Berlin Statement on International Development Training (2008), also need to increase focus on building the demand side and helping policy makers understand the uses and abuses of both monitoring and evaluation. Capacity-building organizations should be training and developing M&E competencies for those who will be primary users of M&E information, requesting monitoring data or evaluations and/or contracting out for evaluations and managing them. Parliamentarians are a key group that needs to understand the uses and abuses of M&E data. It is important to work with users of evaluation results so that they will have a basic understanding of what the "M" and the "E" can tell them, how each can be used to help them do a better job and make policy and practice effective.
7. *Identify complex interrelationships in developing theories of change.* A substantive issue with implications for M&E systems is the growing recognition that it often takes the combined effort of several departments, agencies and programs to achieve strategic priorities. Complex goals cross-cut not only various agencies and programs, but often the non-governmental sector of interest groups and foundations, as well as subnational governments, and the private sector. Performance systems that assume direct linear relationships between an actor and a result will need to be adjusted to reflect complex interrelationships necessary for desired results. This will involve articulating the theory of change behind policy, program, and project interventions. Hatry (2008) indicates a form of matrix development may be needed that would identify the key outcomes sought and each major program or agency (ministry) expected to significantly affect each of the outcomes. Then the individual roles and responsibilities of each program or agency can be delineated and tracked, along with partnership performance. Evaluation would obtain the reasons for unexpected outcomes, especially poor ones. Increasingly, monitoring and evaluation need to change the paradigm to recognize these more complex interrelationships behind an intended change through an intervention.
8. *M&E knowledge banks.* In the context of this chapter, more and more evaluations, disconnected and discrete, are unlikely to produce new and relevant knowledge that is helpful to decision makers, particularly in times of economic crises. As discussed by Furubo (2006), where numerous studies exist to demonstrate the weakness or side effects of a policy, the accumulation of knowledge can lead to a reassessment and a shift in policy. A crisis can provide a window of opportunity, but only if the streams of knowledge exist. Several efforts are ongoing in this regard, including two that are particularly well-known: the Cochrane Collaboration, www.cochrane.org, an international not-for-profit and independent organization, dedicated to providing accurate information about the effects of healthcare interventions; and the Campbell Collaboration, www.campbellcollaboration.org, an international research network that produces systematic reviews on effects of social interventions. Thinking about evaluative systems as about streams, rather than studies, seems a necessary paradigm shift if monitoring and evaluation are to have a seat at the policy table.

Summing Up

Just as one size never fits all, a cookie cutter M&E system is highly unlikely to work. But to be effective, M&E systems must be able to provide timely, relevant, and accurate information to decision makers. Those now toiling in the fields know that such systems can take years to develop and refine. An economic crisis may be a spur, but there are many sources of impetus. Whether born out of crisis or desire for increased transparency, sound data are needed from integrated systems of monitoring and evaluation that take into account capacity building and incentives to use the information. Even with all the basics done well, new strategies will be needed to elevate use of M&E to the highest levels.

The message is thus to embrace evaluation in these turbulent times, but embrace it as part of an integrated system of monitoring and evaluation for effective and efficient use of public funds in turbulent economic times and in good economic times.

References

Berlin Statement on International Development Training. 2008. Final Declaration of the High Level Retreat on the Effectiveness of International Development Evaluation Training. Berlin, June 4-5.

Chelimsky, Eleanor. 1987. What have we learned about the politics of program evaluation? *Educational Evaluation and Policy Analysis.* 9: 199-213.

Furubo, Jan-Eric. 2006. Why evaluations sometimes can't be used—and why they shouldn't. In *From studies to streams: Managing evaluative systems*, ed. Ray C. Rist and Nicoletta Stame, 147-168. New Brunswick: Transaction Publishers.

Gilore, John B., and David E. Lewis. 2006. Does performance budgeting work? An examination of the Office of Management and Budget's PART scores. *Public Administration Review*. September/October. 742-752.

Gray, Andrew, and Bill Jenkins.1989. Policy evaluation in a time of fiscal stress: Some reflections from British experience. *Knowledge, Technology & Policy* 2, no. 4, pp. 20-30.

Hatry, Harry H. 2008. Governing for results: Improving Federal Government performance and accountability. Suggestions for the New Federal Administration. The Urban Institute. November, Washington, DC.

Independent Evaluation Group. (2008). Lessons from World Bank Group responses to past financial crises. *Evaluation Brief 6*, Washington, DC: The World Bank.

King , Jean, Robin Lin Miller, and Melvin Mark. 2009. The Oral History Project Team. The oral history of evaluation: The professional development of Eleanor Chelimsky. *American Journal of Evaluation*. 30, June, pp. 232-244.

Kusek, Jody Zall, and Ray C. Rist. 2004. *Ten steps to a results-based monitoring and evaluation system.* Washington, DC: World Bank.

Lin, Justin Yifu. 2009. Policy responses to the global economic crisis. *Development Outreach*, World Bank Institute, December, pp. 29-33.

Mackay, Keith. 2007. *How to build M&E systems to support better government*. Washington, DC: Independent Evaluation Group, World Bank.

McCormack, Lee, and Bruce Stacey. No year. International foundations for performance budgeting: The case of the Government in Canada. PowerPoint Presentation. June 9 – 10, Mexico City, Mexico.

Morra Imas, Linda G. 2010. The movement for global competencies for development evaluators. A paper presented at the annual meeting of the Canadian Evaluation Society. May 3, Victoria, British Columbia.

Morra Imas, Linda G., and Ray C. Rist. 2009. *The road to results: Designing and conducting effective development evaluations*. Washington, DC: World Bank.

Newcomer, Kathryn, and F. Stevens Redburn. 2008. Achieving real improvement in federal policy and program outcomes: The next frontier. Paper by the National Academy of Public Administration and George Washington University, October, Washington, DC.

OECD, Development Assistance Committee. 2010. Quality Standards for Development Evaluation. DAC Guidelines and Reference Series.

Patton, Michael Quin. 2008. *Utilization-Focused Evaluation*. 4th Edition. Thousand Oaks, CA: Sage Publications.

Pfeiffer, John. 2007. Performance budgeting and performance management in the U.S. Government: Lessons from the PART initiative. Presentation at a World Bank Brown Bag seminar, June 6, in Washington, DC.

Picciotto, Robert. 2005. “Use of evaluation findings to improve development effectiveness”: Panel Discussion. In *Evaluating Development Effectiveness*, ed. George Keith Pitman, Osvaldo N. Feinstein and Gregory K. Ingram, 338. World Bank Series on Evaluation and Development 7, New Jersey: Transactions Publishers.

Richards, Gregory and Swee Goh. 2009. Why is Performance Management So Difficult to Implement? http://www.networkedgovernment.ca/print.asp?pid=917.

Rist, Ray C. 2006. “The ‘E’ in monitoring and Evaluation—using evaluative knowledge to support a results-based management system.” In *From studies to streams: Managing evaluative systems*, ed. Ray C. Rist and Nicoletta Stame, 3-22. New Brunswick: Transaction Publishers.

United States General Accounting Office. 1987. Federal evaluation: Fewer units, reduced resources, different studies from 1980. Washington, DC. GAO-PEMD-87-9.

United States General Accounting Office. 2004. Performance budgeting: OMB’s performance rating tool presents opportunities and challenges for evaluating program performance. Testimony of Paul L. Posner before the Subcommittee on Environment, Technology, and Standards, Committee on Science, House of Representatives, March 11, in Washington, DC. GAO-04-550T.

Weiss, Carol H. 1999. The interface between evaluation and public policy. *Evaluation* 5, no. 4, 468-486.