

CES-ONTARIO CHAPTER CONFERENCE, TORONTO

OCTOBER 1, 2009

“Embracing Evaluation in Tough Economic Times”

KEYNOTE ADDRESS BY LINDA G. MORRA IMAS

Good morning everyone. It is indeed a pleasure to be addressing a group of evaluators today to talk about evaluation in this tough and lingering period of global financial crisis.

Let me start with a quote from Robert Picciotto, former Director General, Evaluation of the World Bank and my former boss, who concludes that the current financial crisis can be traced to “inept evaluations of mortgage loan applications and faulty risk valuations of exotic financial instruments.”ⁱ He then puts forth the interesting hypothesis that had rating agencies been staffed by independent and competent evaluators, we would not be in the financial mess we are in today. I wish we could test this intriguing hypothesis!

Today it is clear that the data to alert us of the coming global crisis of 2008 were available. And while there were a few economists who in fact did predict it, their voices were not heard and their information was not used. This introduction is the lead-in to what is my theme today—

1. economic crises sometimes can be an impetus for evaluation, but there are other factors that come into play;
2. often demand is for performance monitoring systems that can be linked to budgets; but
3. effective response to crises requires sound data from *integrated* systems of monitoring and evaluation;
4. integrated M&E systems must take into account incentives *to use* the data; but
5. a decade of lessons learned on promoting evaluation use points to *need for new strategies*.

Let’s start with the question of whether economic crises frequently increase demand for evaluation. In a global financial meltdown, the first reactions, evaluatively speaking, seem kneejerk ones. How did we get into this mess? Where did the money go? Who or what is to blame? How could it have been avoided? I like to make an analogy to my 21-year old daughter, now out of school, on her own, and trying to manage her finances and having her own financial meltdowns. Sound familiar to any of you? There I am with her, instantly turned into *my* mother, yelling—how could this happen? Where did the money go? How could you have blown so much money without anything to show for it? The evaluative post-mortems, as with my daughter, are a temporary flurry of unproductive activity.

We move fairly rapidly to the second stage, which is: OK, given that we are in this mess, 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 this has happened? For my daughter, the only lesson learned from the past seems to be my spouse and I will have to come up with another personal bailout package. But within our organizations, we may get a more productive flurry of evaluative activity 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. Some of you may have done similar reviews for your organizations.

But isn't there more sustained interest in and promotion of evaluation in tough economic times? There is for some. For the U.S. Government Accountability Office, the GAO, for example, the current crisis must seem like a bull market does to investors. The American Recovery and Investment Act, passed by Congress and signed into law, 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 use of stimulus funds; reviewing and commenting on reports on 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 certainly major challenges are measures of jobs created and jobs preserved. But that is a topic all by itself.

So what about the evaluators like you and me? Now that I have hit the age of mandatory retirement from the World Bank and joined the ranks of independent consultants, this is a real issue for me. It seems common sense that evaluation would be especially important in times of fiscal crisis whether you work for national or provincial government, a non-profit, or a private university. When resources are scarce and competition for them increases, there should be increased need and interest in knowing which programs are achieving intended results with a high cost-benefit ratio and which are not. Such information should be perceived as helping officials make better decisions about which programs to continue or discontinue or to give funding increases or decreases.

But putting these "shoulds" aside, has this been the case? Let's turn the clock back and go to 1980 in the U.S. where the 1979 energy crisis was followed by a 1980 recession. The economy was suffering from slow growth, high inflation, rising unemployment, and unprecedented interest rates. The states were in fiscal crisis and experiencing serious budget shortfalls and slashing budgets. In this tough economic climate, what happened? Ronald Reagan was elected to his first presidential term and took office in 1981. In taking office, Reagan referred to the fiscal crisis he inherited as "the worst economic mess since the Great Depression." Some of you may recall that Reagan had run 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 use of block grants to state and localities. 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 difficulty?

We're evaluators, so let's look at the data. The Program Evaluation and Methodology Division of the GAO 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 GAO in that there was a shift to internal studies—low cost, short turn-around, and non technical—done for program management. The GAO concluded that these kinds of 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 lean economic times are not necessarily a boon for evaluation. Other factors such as political ideology come into play.

Now, being as how we are evaluators, we cannot take this "N of 1". Looking still at the early 1980s, we turn to the case of Australia and get a different picture.

Much has been written about Australia's warm embrace of evaluation in 1983 when a new government faced a difficult macroeconomic situation. Evaluations were viewed as providing the necessary in-depth, reliable information on the efficiency and effectiveness of government programs.ⁱⁱ Formal portfolio evaluation plans were submitted annually to the minister for finance and 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.

But Australia is not the norm either. More frequently in the face of fiscal crises, rather than embrace evaluation, governments chose to develop performance management systems. Let's take another country: Great Britain. In the early 1980s, Britain also had a tight fiscal policy and concern with administrative costs. Andrew Gray and Bill Jenkins reflect on the British experience in time of crisis in relation to policy evaluation.ⁱⁱⁱ Rather than a growth in evaluation in response to this crisis, the reaction was that Britain moved away from its 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". The current fiscal crisis does not appear to be resulting

in increased use of policy or program evaluations although a change of performance indicators from shorter term to longer term was recently called for by Prime Minister Gordon Brown. “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,” Brown told reporters in Brussels ahead of a meeting of EU leaders. The message was one of tying performance indicators to personnel management compensation to create the right incentives.

In New Zealand, another example, 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.

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. I will not list them as eyes would start to glaze, but suffice it to say that no clear link is evident between the 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 fiscal crises, there was no embrace of any form of evaluation.

Here in Canada, a 1994-95 program review that is credited with solving a budget crisis,^{iv} 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.

It is clear that factors other than fiscal crises can lead to the walk down the aisle with performance monitoring and evaluation. Sometimes it is a government committed to 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 performance monitoring.

When a fiscal crisis does serve as an impetus, it seems that it is the “M” that is emphasized, rather than the “E”, and linked to the budget process. It may be called performance monitoring with budget links or results-based budgeting, to emphasize the focus on results, rather than outcomes. But it is a siren call to governments faced with strong pressures not only to account for the billions they spend, but also to show what the money bought, and what have been the results. And it can help finance units deal with the competing demands on the budget.

Did you hear President Obama emphasize program performance and link it to budget actions at his inaugural address? What a message for his first hour of governance! As an evaluator, I was thrilled. Do you recall his words?

“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.”

At last, a nation could understand what an evaluator does!

But as most of you know, there are issues in the use of performance monitoring and performance budgeting systems. At the heart of the matter is the problem that it is difficult to make decisions from limited performance data that are essentially quantitative program scorecards. They 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 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 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, 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.

The concern is that performance monitoring or 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 sometimes interpreted as if they were the direct and exclusive effects of program interventions. We know that we need evaluation 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.

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 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,^v 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,^{vi} and need for disaggregated outcome data by geographic and demographic characteristics and by federal government organizational unit (e.g. region, facility, office)^{vii}.

The largest issue, however, has been use. Executive branch management, funding, and authorization decisions are not regularly based on PART. And in the Congress references to PART have been rare; PART has not been the basis for legislative action.

Now this situation is not unique to PART. Gilmour and Lewis conclude from a 2006 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.^{viii} Richards and Goh point to a recent survey here in Canada of 117 deputy ministers and chief administrative officers that found that only 17 percent used performance measures to make key decisions.^{ix} 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.

But the message I want to give you is this one:

Embracing results-based monitoring systems or results-based budgeting systems may be passionate actions for organizations, especially in times of fiscal crisis, but perhaps short-lived flames unless simultaneously a way is found to embrace evaluation for a lasting relationship in good times and in bad times.

Results-Based Monitoring or Management 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 RBM 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. RBM 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 take time and are expensive. They have little purpose if not used. And if done as a matter of routine on all programs, will they 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 for the demise of Australia's system was the lack of performance information to facilitate the planning and conduct of evaluations.

So where are the examples of monitoring and evaluation 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, there are examples 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 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.

John Pfeiffer from OMB indicated at a World Bank session on PART in June 2007^x that they 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 a lot 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. Again the literature on increasing use of evaluation findings is vast and the main points well-known from building early ownership of the evaluation to report writing (see, for example, Eleanor Chelimsky,^{xi} Carol Weiss,^{xii} Michael Patton,^{xiii} and Linda Morra Imas and Ray Rist^{xiv}). Andrew Gray and Bill Jenkins^{xv} 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. Robert Picciotto believes that "accountability" is critical if to 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."^{xvi}

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 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 my own, co-authored book with Ray Rist, it is the same treatment, but reversed.

So what needs to be done to integrate M&E? In closing let me suggest 7 strategies.

1. 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. And very exciting is that South Africa has just 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, if you will excuse me for another cliché, we cannot let the imperfect be the enemy of the good. We need not only good results-based monitoring systems that are well-linked to evaluation, but we also need a seat at the decision making table to institutionalize routine consideration of M&E data.

2. One way to work toward a seat at the table, raise the visibility of available M&E information, and further routine consideration of M&E data may be the establishing of Monitoring and Evaluation Policy Task Forces by strong evaluation associations. One relevant example is the Evaluation Policy Task Force established by the American Evaluation Association. As described by Eleanor Chelimsky^{xvii} 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.” (pg. 243) As Chelimsky says: “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.” Ideally, we would name such efforts “*monitoring and evaluation policy task forces*” to emphasize the integration of the two, but the idea is that other associations could form M&E Policy Task Forces that work for a seat at the policy-making table.
3. Another new direction 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’s 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 my 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. We might wish that more universities offered comprehensive training for evaluators. But we continue to operate in the main as if no one needs serious training in developing, implementing, and using results-based monitoring systems. It is something that we expect that will get “picked-up” along the way, perhaps with a few hours 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.^{xviii}

4. The same points hold for those developing competencies for those who work in evaluation, including me. Competencies need to be developed for “M&E” and not just the “E”. If we expect complementarity and hope for staff fungibility, then competency programs need to identify the skill sets for both. In the initiative I am leading for IDEAS, we have not given monitoring the billing it deserves, but hopefully we will remedy this!
5. Capacity building statements and positions, for example the recent Berlin Statement on International Development Training^{xix}, also need to focus more on building the demand side and helping policy makers understand the uses and abuses of both monitoring and evaluation. We 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. We need 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.
6. 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. Hatry 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.^{xx} 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, we need to change both our monitoring and evaluation paradigms in recognition of these more complex interrelationships behind change.
7. And while there will never be only one way to build an effective M&E system, guidance in the form of good practice 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 of us involved in M&E will be watching the Canadian Treasury Board’s approach to performance management with great interest. It is, from what I read, 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.

Just as one size never fits all, there will not be a cookie cutter M&E system. But to be effective, M&E systems must be able to provide timely, relevant, and accurate information to decision makers. Those of you toiling in the fields know that such systems can take years to develop and refine. A fiscal 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.

So please do embrace evaluation in these tough times, but embrace it as part of an integrated system of monitoring and evaluation for effective and efficient use of public funds in tough economic times and in good economic times.

Thank you all.

ⁱ Picciotto, Robert. "Where is Development Evaluation Going?" Keynote Address to the IDEAS Global Assembly. South Africa: Johannesburg. 20 March 2009.

ⁱⁱ Mackay, Keith. *How to Build M&E Systems to Support Better Government*. Independent Evaluation Group, World Bank. Washington, DC 2007.

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

^{iv} McCormack, Lee and Bruce Stacey. "International Foundations for Performance Budgeting: The Case of the Government in Canada." PowerPoint Presentation. Mexico City, June 9 – 10 (no year given).

^v Newcomer, Kathryn and F. Stevens Redburn. "Achieving Real Improvement in Federal Policy and Program Outcomes: The Next Frontier." A Paper by the National Academy of Public Administration and George Washington University, October 2008.

^{vi} United States General Accounting Office. "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, 2004. GAO-04-550T.

^{vii} Hatry, Harry H. "Governing for Results: Improving Federal Government Performance and Accountability. Suggestions for the New Federal Administration." The Urban Institute. Washington, DC. November 2008.

^{viii} Gilore, John B. and David E. Lewis. "Does Performance Budgeting Work? An Examination of the Office of Management and Budget's PART scores." *Public Administration Review*. September/October 2006. 742-752.

^{ix} Richards, Gregory and Swee Goh. "Why is Performance Management So Difficult to Implement?" <http://www.networkedgovernment.ca/print.asp?pid=917>. April 2009.

^x Pfeiffer, John. "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, 2007.

^{xi} Chelimsky, Eleanor. what have we learned about the politics of program evaluation? *Educational Evaluation and Policy Analysis*. 1987; 9: 199-213.

^{xii} Weiss, Carol H. "The Interface Between Evaluation and Public Policy." *Evaluation*, Vol. 5, No. 4, 468-486, 1999.

^{xiii} Patton, Michael Quin. *Utilization-Focused Evaluation*. 4th Edition. Sage Pub., Thousand Oaks: CA. 2008.

^{xiv} Morra Imas, Linda G. and Rist, Ray C. *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank. Washington:DC. 2009.

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

^{xvi} Picciotto, Robert. "Use of Evaluation Findings to Improve Development Effectiveness: Panel Discussion", in *Evaluating Development Effectiveness*, George Keith Pitman, Osvaldo N. Feinstein and Gregory K. Ingram (eds), World Bank Series on Evaluation and Development, Volume 7, New Jersey: Transactions Publishers. 2005. Page 348.

^{xvii} The Oral History Project Team (Jean King, Robin Lin Miller, and Melvin Mark). "The Oral History of Evaluation: The Professional Development of Eleanor Chelimsky." *American Journal of Evaluation*. Volume 30, June 2009. 232-244.

^{xviii} There are "independent" evaluation units who only conduct ex poste evaluations and believe their independence would be compromised if they offered any guidance on monitoring indicators early in the program or project design stage. Most evaluation is not conducted, however, under such constraint

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

^{xx} Hatry, Harry H. "Governing for Results: Improving Federal Government Performance and Accountability. Suggestions for the New Federal Administration." The Urban Institute. Washington, DC. November 2008.