PART ONE – THE BIG PICTURE

Not everyone reads publications sequentially from front to back. Most of us dip and dive in our reading. Despite this there is always great debate between the editors about which article will start a publication. Nearly always. Not here. As soon as the early drafts of this article arrived, it was clear that it formed a superb scene setter. It's an excellent description of the key practical elements of successful evaluations, based on an open and honest reflection on a long career. It also contains sage advice for those seeking to promote evaluation practice as a standard and uncontroversial part of program management. The observations are both timely and timeless.

The pillars of public program evaluation

George F. Grob

In this article, I offer perspectives on evaluation gained from a life-long career in preparing and using evaluations in the U.S. Federal Government, mostly at the Department of Health and Human Services. I wish from the outset to acknowledge the narrowness of this perspective. It reflects only a portion of the worldwide practice of evaluation. The evaluators of South Asia can learn much from their colleagues around the globe, as many nations have taken up evaluation as a way of helping policy makers and program managers make informed choices about developing and implementing public programs.

Having said that, the incorporation of evaluation into the workings of U.S. Federal domestic agencies, particularly in the fields of education, public health, and social services, has been a deliberate enterprise with concrete origins traceable as far back as the late 1960's. Those of us who started in those early years had little to go on by way of formal evaluation training, standards, and precedence. We figured a few things out and made a lot of mistakes. The evaluators of South Asia need not retrace our confusion and errors. We ourselves still have much to explore in the world of evaluation. However, we can at least share what we have already learned, including the "don't ever try that again" lists we have compiled along the way. It is in that spirit that I offer not a visionary dream of the future but a practical blueprint for a functioning evaluation system that the nations of South Asia might want to use as a reference point in building their own.

The article is written in two parts. In Part One, I will describe the fundamental pillars of a national evaluation system. In the process I will comment on the feasibility of and resources for the further development of evaluation in South Asian countries. In Part Two I will reflect on my own experience as an evaluator and suggest ways to move quickly to implement national evaluation programs based on the lessons I've learned so far.

Part One
The Pillars

The pillars upon which a well functioning evaluation edifice may be solidly built include:
1. understanding evaluation types and methods and how policy makers and managers use them
2. appreciating the relationships among evaluation and related professions
3. establishment of government agencies that evaluate public programs
4. freedom and encouragement for growth of non-government evaluation organizations,
5. evaluation education and training programs, and
6. professional standards.

PILLAR 1
Understanding evaluation types and methods and how decision makers use them
As a common noun, the term "evaluation" is fairly easy to understand. People evaluate things all the time. Formal evaluation, as a profession and administrative function, is a disciplined and methodical approach to this natural and common sense activity. By using formalized methods and principles, professional evaluators can help policy makers, program managers, and other interested parties obtain more evaluative information and more reliable and trustworthy data than might otherwise be available to them.

Formal evaluation encompasses a broad range of types and methods, depending on the objective of an evaluation study. Some are highly focused, others are broad in scope. Some can be done quickly, even within a matter of days or weeks; other might take one or more years to do. They can focus on administrative compliance, efficiency, outputs, outcomes, or impact. Some are highly focused while others are exploratory.

New methods and analytic approaches are being developed all the time. Innovations over the years have included focus groups, inspections, quantitative methods, qualitative methods, program logic models, participatory evaluation, randomized assignment groups, and systems analysis, to name just a few. While these phrases may seem complex and intimidating, they are simply formalized approaches that build on our natural inclination to assess important matters.

One way to understand the diversity of purposes and methods upon which the professional field of evaluation is based is to think of the natural course of public programs. Many government programs have been around a long time, but all of them were once new, and they probably went through some fairly typical stages as they matured. Let us suppose for a moment that an innovative method of delivering health care in a rural area show promise for ease of administration, efficiency, and impact. Let's say that a government decides to start this program and build on it in the future should it prove beneficial. Here is a bird's-eye, time-compressed view of how a formal evaluation program can help promote the success of such a program or support difficult decision making about it.

In the beginning, public officials might be concerned as to whether the program is well defined and understood. Evaluators can help public officials sharpen their understanding of immediate, intermediate, and long term goals and to advise them on how well the administrative structure is conducive to obtaining the intended results. This independent and professional advice about the structure, resources, services, beneficiaries, intended short term outcomes, and the potential long term impact on public health is an example of the application of program logic models.

Once underway, program administrators might want to know if the program is being implemented as intended and if any unexpected barriers have arisen to prevent or delay implementation or whether any unintended untoward effects have surfaced. Rapid highly focused field inspections, as well as quick surveys of "on-the-ground" administrators and beneficiaries, can bring back that kind of information to decision makers so they can make immediate and mid-course corrections.

As the program matures, public officials and other interested parties may want independent, professional reviews of a broad spectrum of issues related to compliance with official administrative policy, cost, nature and extent of services provided, number and kind of beneficiaries being served, early indications of the beneficial impact on individuals' health, etc. All of these questions are amenable to fairly quick evaluation using established methods.

Over time, public officials, interest groups, and citizens will want to know how well the program is performing overall, whether it is delivering the health impacts intended, how well this program compares to other approaches that have already been used, how well it compares to other approaches used in other regions or countries, and whether there are unpredicted side effects both good and harmful. Evaluation studies can also be used to provide some answers to these broader questions. However, these studies may be more complicated and take more time

and resources to perform than some of the studies mentioned earlier.

During the entire lifetime of the program, from start-up through early implementation and maturity, it may be desirable to have systems that routinely gather information about resources, services, beneficiaries, goals, outputs, outcomes, and impacts. This can facilitate the ongoing monitoring and evaluation of the program in addition to the kind of periodic studies discussed above.

All of these situations reflect the need for a variety of evaluation approaches and methods. Professional evaluators can provide practical, independent advice how to construct and carry out any of these assessment activities. Here are some examples of the kinds of studies that were performed in the Department of Health and Human Services where I worked.

In 1997 legislation had been passed to facilitate enrollment in health insurance for low income children (the U.S. does not have universal publicly funded healthcare). At that time, some 12 million children were without insurance. The new program, called the Children's Health Insurance program, provided $24 billion over five years to develop new insurance programs and to streamline enrollment. Since then, the Office of Evaluation and Inspections of the Office of Inspector General has routinely conducted studies to monitor how this program is being implemented and whether it is meeting its goals. The first study was started within the first year of the program's implementation. Entitled "States Application and Enrollment Processes: An Early report from the front Lines," (http://oig.hhs.gov/oei/reports/oei-05-98-00310.pdf), the study gathered information in person and by telephone from program officials and front line workers and obtained copies of forms used to enroll children. The study found that some States were using innovative ways to enroll the children quickly, including making application forms available in many sites, offering them in several languages, and offering toll-free telephone lines to facilitate information sharing and availability. However, it also discovered significant barriers such as citizenship dilemmas and complexities arising from rules of multiple programs for which the children might be eligible. It made recommendations to improve the readability and simplicity of the enrollment forms and to streamline the enrollment process based on the promising approaches of some of the States..

Subsequent studies of this program focused on parents' perspectives of the administrative processes and hurdles for renewing their children's insurance coverage (http://oig.hhs.gov/oei/reports/oei-06-01-00370.pdf); how States assess their own progress in meeting their enrollment goals (http://oig.hhs.gov/oei/reports/oei-05-00-00240.pdf); and problems they encountered in measuring progress towards reducing the number of uninsured children (http://oig.hhs.gov/oei/reports/oei-05-03-00280.pdf).

Other examples of evaluation studies intended to help program managers and the beneficiaries they served include a series of reports on children's use of health care services while in foster care, e.g., one relating to children in Kansas(http://oig.hhs.gov/oei/reports/oei-07-00-00640.pdf); physician participation in the Vaccines for Children Program (http://oig.hhs.gov/oei/reports/oei-04-93-00320.pdf); the role of the Federal Maternal and child Health Bureau in preschool immunizations (http://oig.hhs.gov/oei/reports/oei-06-91-01180.pdf); and children's dental services under Medicaid (http://oig.hhs.gov/oei/reports/oei-09-93-00240.pdf).

Over the last 20 years, the Office of Evaluation and Inspections has issued over 1000 reports of interest to policy makers in the Department and the Congress. These studies covered a broad range of programs and topics, including Medicare and Medicaid, prescription drugs, nursing home care, home health, hospice care, mental health services, food safety, dialysis, dietary supplements, human research subjects, transplant organs, hospital accreditation, Social Security, disability, rehabilitation, foster care, bioterrorism, and child support enforcement. They addressed the management, efficiency, and effectiveness of these programs as well as their vulnerability to fraud and abuse. The studies, along with associated audits and investigations, are credited with saving billions of dollars and achieving favorable impacts for program

beneficiaries and U.S. citizens and in significant improvements in the quality of services and the effectiveness of programs. To review thee reports visit the web at http://oig.hhs.gov/oei/atod.html.

PILLAR 2
Appreciating the relationships among evaluation and related professions

As natural as evaluation is, and perhaps for that very reason, some people are mystified by the claim that it is a profession. Almost everyone recognizes that auditors, social scientists, statisticians, mathematicians, psychologists, economists, and public administrators are members of respected professions that one can aspire to, train for, and practice. But it is only recently (over a few decades) that evaluation has emerged as a distinct profession. To a significant extent the maturation of this profession is still unfolding.

The distinguishing characteristic of evaluation is its focus on the value of programs, usually government programs or those of non-profit foundations that seek to improve human conditions. Having said that, evaluation does have much in common with several related professions and fields of study. In fact it has derived many of its methods from related fields, including audit, economics, social science, psychology, logic, statistics, and mathematics. Furthermore, it is a prominent aspect of some programmatic fields of endeavor such as education, public health, social services, public administration, environment, and public policy. Many people with advanced degrees in these areas include evaluation as one part, sometimes the major focus, of their professional practice.

Upon reflection, the overlap of evaluation with other professional fields proves beneficial in several significant ways. It means that the burden of evaluation does not fall exclusively on those who claim to be evaluators. In fact, an important first step in evaluating any public program is to find out what the public literature of other professions has to offer about the program. It makes no sense for evaluators to duplicate work already done. Another benefit is the growth of understanding as evaluators and other professionals begin to interact and work together, thus triggering the well known process of creativity and insight that is the predictable outcome of such exchanges.

There is another advantage well worth emphasizing here. Enlarging the pool of evaluators, especially those in the early stages of their careers, can easily be facilitated by hiring from the pools of other professions. As noted earlier, many who have graduated with degrees in education, public health, social services, and the environment, to name just a few fields of study, are quite interested in evaluation and may be looking for jobs that allow them to use their training in the service of the public good through evaluation. The same can be said of mathematicians, statisticians, auditors, psychologists, and other with functional specialties. All of these aspiring professionals come well equipped to conduct evaluation studies, and their transition to the field of evaluation is an easy and natural process.

PILLAR 3
Establishment of government agencies that evaluate public programs.

The U.S. Government now contains several significant offices that perform evaluations for policy makers. The single largest evaluation agency in the United States is the Government Accountability Office (GAO). It has some 2,500 employees, most of whom perform evaluations at the request of the U.S. Congress. Not too long ago, it was known as the General Accounting Office (GAO). The recent name change reflects the gradual change of focus in its work from audits to evaluations. Some Offices of Inspectors General (one of which is situated in each of the Federal executive departments and agencies) also contain units that perform program inspections and evaluations similar to those conducted by GAO. Furthermore, the legislatures of virtually every State in the U.S. also have reporting to them an independent agency modeled

after the GAO.

Many federal agencies have one or more evaluation offices whose full time responsibility is the evaluation of agency programs. For example, the U.S. Department of Health and Human Services has an Assistant Secretary for Planning and Evaluation who reports directly to the Secretary and whose office helps in planning and coordinating evaluation activities across the Department. There are discreet evaluation offices within many of its constituent agencies, including the Centers for Disease Control and Prevention, the National Institutes of Health, the Substance Abuse and Mental health Services Administration, the Administration for Children, Youth, and Families, and the Office of Inspector General.

The existence of these evaluation offices illustrates both the need for, and the natural outcome of, embracing evaluation as an integral part of public administration. Eventually, any nation that wishes to systematically use formal evaluations to promote program success will probably find it necessary to establish or support permanent agencies to perform evaluation studies.

PILLAR 4
Freedom and encouragement for growth of non-government evaluation organizations
Evaluation can be a powerful force for social betterment if its influence is absorbed and assimilated throughout society. It is not sufficient that governments produce evaluation studies. Evaluation's impact will best be felt if it is studied and performed in institutions of higher learning and if non-government public interest groups have the ability to obtain professionally prepared evaluation studies on topics of importance to them. If it catches on, there may naturally arise for-profit businesses capable of producing sophisticated studies while retaining their own independence from the inappropriate influence of those who pay them for doing so. Similarly, the non-government public interest groups themselves may acquire the ability to perform evaluations or to sponsor their performance by other non-profit organizations. One way for government agencies to promote and support the growth of non-government evaluation sources is to contract out the evaluation work. The subject of non-government evaluations is a broad topic beyond the scope of this paper. It is mentioned here, however, to introduce this notion early in the minds of those wishing to influence the development of a national evaluation system.

PILLAR 5
Evaluation education and training programs
The introductory section portrayed evaluation as an emerging profession and linked it to related predecessor professions. Nowhere is this more evident as in the field of evaluation education and training. When people of my age first entered the field of evaluation, there were, for all practical purposes, no sources of formal evaluation education. Most of us started our careers in one or more of the allied professions discussed earlier and gradually began learning from one another, sharing our experiences first informally, then through professional evaluation associations. The largest of these in the United States, the American Evaluation Association (AEA) (http://www.eval.org/), recently celebrated its 20th anniversary.

Initially, there was a strong interest in the field of evaluating education programs. This had the salutary effect of enticing evaluators to work in university settings, an environment conducive to inquiry into methods of thinking. That led to college and university courses in evaluation itself. Gradually this idea spread, and slowly but surely evaluation has become a core subject in the curricula of U.S. college and university schools of education, public health, social services, and other such disciplines. The American Evaluation Association lists on its website some 45 universities that offer doctorates, masters degrees, or professional certificates focusing on evaluation (http://www.eval.org/Training/university_programs.asp). This includes

institutions of higher learning in Australia, Canada, Denmark, Israel, New Zealand, Puerto Rico, and South Africa.

It is worth noting however, that in almost all U.S. cases evaluation is a subject of interest in a school whose primary focus is education, social services, scientific methods, or other such academic fields. Evaluation in the U.S. largely remains in the embrace of allied fields and professions as discussed in the very beginning of this paper. The advantages of this situation are that there is now no lack of university level training in evaluation in the U.S. and evaluation is being integrated into many professions and fields of study where its influence will certainly grow.

A disadvantage, though, is a scarcity of training in the general disciplines of evaluation. Such education and training is crucial to reach individuals whose professional evaluation practice will not be in the fields where it is already well established and to draw highly talented generalists with good command of writing, analysis, and scientific methods into evaluation practice. Gradually this need is being met by the international, national and regional evaluation assocations, societies and networks. Over twenty national professional evaluation associations have been formed, including evaluation societies in Malaysia (http://www.mes.org.my/), Pakistan and Sri Lanka (http://www.nsf.ac.lk/sleva/). In addition, international evaluation associations cover South Asia (ENSA), Africa, Europe, Latin America, and Russia and the Newly Independent States. All of them support conferences and publish a variety of studies and handbooks for their members. This publication explores the work of two national evaluation societies. A convenient list of national and international societies can be found at http://www.eval.org/Resources/ProfessionalGroups.asp.

Other sources of evaluation training come in the form of published books and proprietary training courses, all now more abundant than in earlier years. Many of them provide expert professional training on the fundamental topics of evaluation – theories of evaluation, statistics, sampling, qualitative and quantitative methods, and many more topics.

PILLAR 6
Professional standards

Well established professional organizations adopt standards of ethics and professional practice for their members. The American Evaluation Association (AEA) has done just that, and has gradually refined and strengthened them over the years. They call for systematic inquiry of the highest technical standards; collective competence of evaluation teams practicing within the limits of their professional abilities; integrity and honesty; respect for people; and responsiveness to general and public interests and values. The latest version, "Guiding Principles for Evaluation", (http://www.eval.org/Publications/GuidingPrinciples.asp).

The AEA is not the only evaluation organization to develop standards for the practice of evaluation. Evaluation associations of Australia, Canada, France, Germany, Italy, Switzerland and the United Kingdom have all published standards, as have international groups such as the African Evaluation Association, the European Union, and the United Nations Evaluation Group. All of these standards could provide a good starting point for any national or regional group in preparing its own. The European Evaluation Society provides a useful listing of many these international standards on its website (http://www.europeanevaluation.org/?page=756983). Any of these easily could be adapted and "localized" by any group wishing to establish its own set of standards and ethics.

Another useful set of standards are the Program Evaluation Standards of the Joint Committee on Standards for Educational Evaluation (http://www.wmich.edu/evalctr/jc/). This committee is sponsored by numerous U.S. national organizations dedicated to quality education and by the American Evaluation Association. Although prepared in the context of evaluating educational programs, they can easily be adapted to many other subjects. They offer guidance on such subjects as evaluator credibility; values underlying an evaluation; identification of

stakeholders; clarity, timelines, and dissemination of reports; planning for impact; feasibility and cost effectiveness of the evaluation effort; rights of human subjects; respect for the human dignity and worth of participants in the evaluation; conflict of interest; completeness and fairness of the assessment; and the validity and accuracy of methods and documentation.

Other useful standards are available from government organizations that prescribe standard procedures to be followed by their staff. The GAO, previously mentioned, issues standards to be followed by any organization performing audits of Federal government programs. Of interest to evaluators are the sections in the standards that have been added to govern the evaluation work which most of its employees now perform. These new sections refer to evaluations as "Performance Audits". Those sections, as well as those of the general introduction that deal with ethical principles, independence, professional judgment, and competence form a solid basis for the professional work of evaluators of government programs. The January 2007 version can be found at http://www.gao.gov/govaud/d07162g.pdf.

Another example is the standards that are issued by the President's Council on Integrity and Efficiency (PCIE). The PCIE standards govern what are known as "inspections and evaluations", a family of studies that are performed by Federal Government Offices of Inspectors General, mentioned previously. Like the GAO standards, they provide practical guidance for planning studies, quality control, data collections and analysis, evidence, and reporting. They can be found at http://www.ignet.gov/pande/standards/oeistds.pdf.

Like the AEA and Joint Committee standards, the GAO and PCIE standards are particularly suitable for monitoring and evaluating government programs, and, like the AEA and Joint Committee standards, they can easily be adapted to the work of evaluators in other settings. They also have the advantage of getting down into the nuts and bolts of how evaluation studies should be planned, designed, executed, managed, and published.

Part Two

Looking Back and Taking Forward – Reflections On Being An Evaluator

Building the necessary infrastructure for a vibrant culture and practice of public program evaluation in South Asia may at first sound somewhat daunting. But those interested in seeing the fulfillment of the idea do not have to start from scratch. They can take advantage of the resources that already exist. The ease of international communication through cellular phone technology, the Internet, fax machines, and international travel put almost everything that is both known and published about evaluation within fairly easy reach of any professional group on the globe.

The proof of feasibility and the hope of success are demonstrated by the very publication in which this article appears. It seeks to draw attention to the benefits of public evaluations – across all of South Asia and to a broad readership including not only evaluators but also professionals and interested parties throughout both government and civil sectors of society. In addition, steps have already been taken to attract university programs from other countries to begin the process of education in this field. With such energy, commitment, and vision, it is hard to believe that this enterprise will be anything but successful.

In the context of these promising events, I was asked to offer some ideas based on my own experience about how the infusion of professional evaluation practice in public policy making and program management might be expedited in South Asia. I will approach this from two perspectives – the bottom and the top. By the "bottom" I mean the professional practices of individual evaluators or evaluation offices; by the "top" I mean those broad and sweeping forces that, if harnessed, can facilitate the adoption of formal evaluation programs on a large scale. The "bottom" is what we learned by doing in the past. The "top" is what we see evolving across many nations.

Looking Back

The insights that I will discuss from the "bottom" come from my own experiences and those of my immediate colleagues entering the practice of evaluation over the last thirty or so years. As I mentioned before, we had little to go on when we started, but we did our best. Gradually, professionals younger and better trained than us came along and demonstrated the value of more formalized aspects of our profession by their own effective professional practice of evaluation. But we could show them a thing or two that we had learned the hard way. Please allow me to do a bit of that here. I will start with some things I think we did right. Then I will move to some of the lessons learned as we gradually refined our craft.

- **Effective Writing.** First and foremost we recognized and emphasized the importance of clear, simple writing as the best way to deliver the results of evaluations. We called it "Write to the Point" and required all of our staff to be formally trained in its principles. We found that we could get policy makers to understand us and take action based on our work if we spoke our truth in plain language. We tried to avoid long words and complicated sentences. Some of us talked about "The Mom Test" – the principle that our own mothers had to be able to understand what we were talking about and to be inspired to take action based on what we wrote. If I were told to chop this entire article down to one paragraph that I wouldn't want any aspiring evaluator to forget, it would be this one.

- **Working in Teams.** A simple rule we learned early was "Never Work Alone". By definition, every study we ever did was dealt with problems that no one had quite solved as of yet. Most policy makers and program managers are bright and dedicated people. They want their programs to work. If they knew everything about how to run their programs they would not need our studies. We found that we could benefit from many points of view – from program proponents and critics, subject matter specialists, methodologists, administrators, statisticians, policy experts. We did our best to draw them all into our studies, giving them a chance to weigh in at the beginning, during, and after the study process. We are grateful for their many insights obtained over the years. Their real impact was on program beneficiaries and administrators who learned something about how to improve public programs from our studies.

- **Sticking to the Evidence.** We also learned to have a healthy distrust of our own thinking and to avoid speaking beyond the evidence we had collected. Perhaps it was just wishful thinking on our part, but we wanted to make suggestions on how to make things better. However, solid proof is a valued commodity in the world of public policy making. We therefore began to study the nature of proof and evidence and to follow the principles we learned with increasing rigor. To keep ourselves honest, we invited cold readers to review our work and challenge our conclusions and recommendations. This fire, though sometimes stinging, produced more gold. As our work became more publicly visible, we sharpened up our delivery of evidence. We began to see our standard product as a short report and a box of evidence. For evaluation, as for professions involved with systematic inquiry, public scrutiny is the most effective and unforgiving quality assurance program.

- **Using Multiple Methods.** Closely aligned to the previous point was our gradual realization that everybody had their own way of thinking. We foud that evidence that was compelling for one reader was derided by another. The solution, we learned, was to always obtain more than one kind of evidence. Sometimes we would use three or more approaches to study a problem – examination of billing records, reviews of administrative files, beneficiary surveys, focus groups, statistical analysis. Usually the most effective evidence was a combination of "hard data" (one or more forms of quantitative analysis) and the systematic collection of the perspectives of affected parties (one of many kinds of qualitative analysis). Some of readers told us that

numbers always lie, so we gave them the perspectives of real people. Others believed the opposite, that "people always say that;" so for them the numbers worked better. We found that the combination of the two was powerfully persuasive.

- **Investing in Designs.** Early in our work, we were anxious to get into the field and learn first hand what was happening to programs. Later we discovered that we had not been as disciplined as we needed to be in deciding what kind of data to collect. We wasted time and effort and returned with too much confusing information. Although it slowed us down at first, we eventually found that our studies could be done faster and far more effectively if we took more time up front to clarify what issues we would focus on, what kind of evidence we would gather, and who our readers would be. We developed a standard "design" module to follow in the early stages of each study. That helped a lot.
- **Strategic Thinking.** Eventually we realized that simply doing studies was not enough. Initially, we invited our staff to voluntarily suggest topics for study. Later we learned to be much more systematic. We first picked broad areas of interest – e.g., child welfare, health care for the poor, nursing homes, home health – and to invest in learning about these areas and what other experts and scholars had already learned. We started using our studies to fill knowledge gaps. This was far more effective than picking topics just because we were interested in them. It gave us the advantage of building on what a lot of other people had done before us.

Taking Evaluation Forward

What might my own experience suggest for taking evaluation in South Asia, already on the move, further forward? Here are some ideas:

- **Evaluation offices can start producing evaluation studies right off the bat.** There is no need to engage in a prolonged planning and training program before conducting studies. It is possible to do both – do and learn at the same time. Policy makers and program managers will not be impressed with promises about how good things are going to be. As counterintuitive as it may appear, evaluation offices may find that their credibility is enhanced if they begin producing evaluations as soon as possible. Newly established evaluation offices might want to begin with short, highly focused monitoring studies. These are fairly easy to learn and do, but are very appreciated by program managers and national policy makers alike. It will be fairly easy to branch out to more complicated studies. Expert professional help always enhances success no matter how advanced the evaluators are. The methods for managing contracts for such studies is easily learned and implemented.
- **National or regional evaluation education centers can provide the anchors for the steady growth and development of evaluation.** Centers for promoting training and development in evaluation should preferably be located in one or more university settings and with links to similar centers around the world. Some initial discussions of this idea for South Asia have already taken place.
- **Quick-start, evaluation training "boot camps" are also beneficial.** Professional training programs can be made available even outside the university setting. Evaluation training can start quickly by focusing on the essentials – writing, statistics, sampling, surveys, negotiation, evaluation design, and evaluation planning and principles. Such introductory training can help professionals of allied fields and well educated individuals of all kinds who are interested in entering the field of evaluation to quickly produce professional studies that can be useful to pubic decision makers at all levels of government and private philanthropies.
- **Evaluation societies can lead the way.** The evaluators themselves can successfully promote professionally sound and creative approaches to and training for evaluation.

These societies can be modeled after the national and international groups discussed above. South Asian professional evaluation societies that already exist can be particularly helpful here. The new Evaluation Network of South Asia (ENSA) could be an important addition. In any case, support by government agencies and non-government organizations for evaluators to attend professional conferences will pay off in both the short and long run.

- **The wide publication of this publication will spread the news of evaluation and its benefits throughout South Asia's institutions.** This publication and publications like it, telling stories and experiences that are locally relevant can promote the need for and cultivation of evaluation as an important tool of public policy making and program management.

- **Professional standards can be developed by the evaluation societies.** Evaluation standards can be modeled after those used in other nations and by international groups. Issuance of the standards can garner much attention in the media, government, education community, and social interest groups. This will further promote a broad understanding in society and governments of the benefits of formal evaluation programs. And it provides a sometimes necessary safeguard for those undertaking evaluations.

- **Government agencies should be encouraged to provide public funding of evaluation studies.** One way to do this is by setting aside some portion of public program budgets to be used for obtaining independent professional evaluations of the programs.

- **The establishment of independent government evaluation agencies would provide a particularly strong foundation for the use of evaluation in public policy making.** Independent government evaluation units reporting directly to the legislatures and senior government ministries would ensure an ongoing supply of highly professional evaluation studies that are focused on important public policy issues and for these reasons be much likely to be used for public decision making. Such agencies will need sufficient scope, funding, authority, and independence to be helpful and trusted.

I hope these observations are helpful to the many people working so hard to build up the capacity for public program evaluation in South Asia.

At first glance you might doubt the relevance of this article to projects and evaluations in South Asia. It deals primarily with the "translation" of projects developed in very controlled circumstances to the "real world". Yet, whilst controlled experiments like this are rare in South Asia, pilot projects are not. Evaluations designed to help "mainstream" pilot projects are more difficult than they seem. Evaluations in South Asia are also called on to help translate the experience of a particular project in a particular area to more widespread application.

So the issues and suggestions covered in this article are highly appropriate to South Asia. Doubly so. Firstly the issues of translating "research" to "practice" are pretty much the same as mainstreaming pilot projects or transferring projects from Area A to Area B. Secondly, and no less important, the article highlights some largely undiscussed issues that are very relevant to the current pressure by some donors to fund only "experimental designs" (e.g. random controlled trials) or "quasi-experimental" evaluation designs.

Key Evaluation Issues in Facilitating Translation of Research to Practice and Policy

Russell E. Glasgow, Ph.D

Summary

This article summarizes critical evaluation needs, challenges, and lessons learned in translational research. Evaluation can play a key role in enhancing successful application of research-based programs and tools and lead to efficient and sustainable programs.1 Discussion centers on what is unique about evaluating programs and policies for dissemination (or potential for dissemination). Central issues reviewed include: the importance of context and local issues; robustness and applicability across a wide range of settings; multiple levels of evaluation; implementation fidelity vs. customization; who participates; and characteristics of "success" at each stage of recruitment, delivery, and outcome. The use of both qualitative and quantitative is especially important and the primary redirection that is needed is to focus on questions of decision makers and potential adoptees rather than the research community.

What is Translational Research?

Almost everyone agrees that there is a serious problem concerning the slow and incomplete transfer of research findings into practice (Institute of Medicine & Committee on Quality Health Care in America, 2003). This seems to be true across diverse content areas, countries, and areas of specialization (McGlynn, Asch et al., 2003). There are multiple and interacting reasons for the present situation including resources, training, reimbursement and other policies, priorities, vested interests, and political will (Kingdon, 1995). *This article focuses on what evaluation science can do to facilitate successful transfer of research to real-world practice and policy by helping to plan and evaluate programs and policies that have broad reach, increased impact, can be consistently implemented, are more efficient, and sustained.* The article discusses the characteristics of implementation and dissemination research and the related evaluation implications. It illustrates application of a research translation evaluation model (RE-AIM) and concludes with a list of specific translation challenges and recommendations for evaluation.

1   Or the translation of pilot projects to mainstream, or duplicating projects to other settings.

Unique Features of "Translational Research"

The key features of translational research are concerned with either evaluating – or collecting data to estimate results – typical citizens in typical settings receiving interventions delivered by typical staff (Glasgow & Emmons, 2007; Green & Ottosen, 2004). This is in contrast to research that focuses on a subset of persons most likely to benefit and without "confounding factors", the kind of research often conducted in academic settings.

The primary purpose of translation research is to address practical questions that key decision and policy makers are likely to have (e.g., can this program work here; how much will it cost; who can successfully deliver the program?). Theory is important in this type of evaluation, but questions revolve around how the theory is implemented, rather than around more basic theoretical questions.

Those who have performed this type of evaluation work consistently comment on the importance of context. By context, I mean the historical, setting, organizational systems, and resource conditions that surround and may influence the outcome of a program or policy. To capture context adequately often requires relatively comprehensive evaluations, as described in later sections. The programs and policies evaluated are often "complex", multi-component, and multi-level, which present added evaluation challenges. A related feature of translational research is that interventions or policies often evolve over time. Sometimes this is intentional as in rapid cycle quality improvement programs (Berwick, 1996), and other times it is unintentional and due to changes or "drift" in staff, available resources, or priorities.

The characteristics of translational research and evaluations that are intended to inform policy and practice also have important implications for how evaluations are conducted. These implications are discussed below.

Context

When evaluating programs for translation, it is helpful to ask questions that a journalist might (Table 1). A key question that is challenging to answer, but which has enormous implications is "Who Participates – and Who Does Not?" Typically this question is answered only in terms of the "numerator" of the number of participants (e.g., citizens, employees, students) who took part, the number of staff who delivered the program, or the number of settings assessed. By itself, such information is only moderately helpful. Much more useful is to also collect information on the "denominators" of the intended numbers of participants, staff and settings invited to participate; and on the similarities and differences between those who take part and those who do not at each of these levels.

TABLE 1
Key "Journalist" Contextual Questions for Evaluating Translational Programs and Policies

1. WHO attends? (and who does not) – at following levels:
Setting: which communities or organizations (e.g., worksites, medical plans, schools) were approached – how many participated?
Staff: which staff members participated, and
Individual citizens, consumers, end users: How many and what types of people participated?

2. WHAT outcomes are produced? (intended and unintended)
How much change is observed on key dependent variables?
What is the impact on quality of life?
Were any negative impacts produced?

3. WHEN and where will this program work?
What types of settings and staff members are most successful?
What participant characteristics are associated with success?

4. HOW consistently was the program/policy delivered?
Across different program components?
Across staff?
Over time? (Did program change?)

5. HOW long-lasting are the effects?
What was the attrition—at setting, staff, and individual levels, and how did this impact results?
To what extent were participant improvements maintained over time?
Was the program or policy continued/institutionalized, modified (and how) or discontinued?

6. WHY were these results found?
How did change come about (what were the mediators?)
What contextual factors were important?

Most projects can collect numerator and denominator information by simply keeping careful records and reporting this information, as well as the number of and reasons for exclusions at each of the levels in Table 1. Due to confidentiality or logistical issues, it can be more challenging to collect information on characteristics of those who decline to participate. In such cases, a useful fall-back strategy is to rely on existing data sources such as local census data, reports from health departments, or organizational records to compare the characteristics of the target population in that area (e.g., all employees in the work force, all citizens in a neighborhood or community – see www.re-aim.org) to those who participate.

What Outcomes?

The second key question in Table 1 concerns the magnitude and breadth of improvements produced by a program or policy. Many evaluations are limited to a narrow assessment of impact on a pre-identified key outcome. Amount of change on this primary dependent variable is one important aspect of evaluation. Equally important, however, are answers to the related question of impact on quality of life – considered by many to be the ultimate outcome of goal of public services (Kaplan, 2003), and to know if any negative or unanticipated results occurred.

Often program and policy developers have difficulty identifying potential adverse events that might occur as a result of a new program. One of the main ways that programs can have a subtle negative impact is that by focusing efforts on a given area (e.g., education), busy and under-resourced settings may do less in other areas such as immunizations or mental health.

When and Where Will the Program Work?

The third set of questions in Table 1 assesses the breadth of conditions under which a program is successful. At the setting level, this refers to organizational characteristics related to success. For example, are only well-resourced settings that employ a multi-disciplinary team approach able to achieve success? At the individual or consumer level, a key issue is whether results are uniform or differential across recipient characteristics such as race, ethnicity, income, education, gender, age, and literacy levels.

Consistency

The fourth question concerns how consistently programs are delivered across different intervention components, staff, recipients, and time. Consistency of delivery by typical staff is especially key in evaluations conducted in real-world settings because failure to adequately implement a program is a frequent reason for failure (Basch, Sliepcevich et al., 1985). It is important to understand both the extent to which different aspects of a program are delivered as intended, and if there are staff characteristics (e.g., education, profession, experience, similarity to recipients) associated with successful program implementation.

It is also important to track program/policy implementation over time. Intervention delivery patterns can evolve over time, both intentionally and unintentionally. The issue of program "fidelity vs. customization" is currently an active area of investigation and controversy, and is discussed in more detail later.

Sustainability

The fifth question in Table 1 concerns the longevity of programs and their effects at both the setting and individual levels. If an organization or government agency is going to make an investment in a new program or policy and devote the time and resources involved in training, supervision, infrastructure, etc., it wants to have a reasonable expectation that both the program (policy) and its effects will stand up over time.

At the individual level, there are two key evaluation issues related to sustainability. The first is attrition. It is often challenging to track participants over time in mobile societies, but attrition rates can produce misleading conclusions to evaluations. This is especially the case if attrition rates are high, are related to participant characteristics (especially to success), or are differential across program conditions. The other well-known issue is that of maintenance of results over time. Many problem behaviors and societal issues can be modified over a short period of time, but long-term maintenance is a much greater challenge (Orleans, 2000).

At the setting level, the key sustainability question concerns whether the policy or program is continued intact, discontinued entirely, or modified following an initial evaluation period. There are little data about the extent to which organizations adapt, modify, or discontinue programs over time (Glasgow, Klesges et al., 2004); but in my experience, it is rare that a program is continued in exactly the same way it was evaluated during a study. We need to study "evolution" of programs over time to enhance understanding of translation issues.

Understanding Why

The final key issue is to provide information on how and why the pattern of outcomes observed was found. The goal is to understand how the program or policy achieves its effects (or why it did not succeed, or was only effective for a subset of participants). Both quantitative and qualitative approaches are often helpful to elucidate such understandings, which can inform both program refinement and the underlying theory. Baranowksi et al., (Baranowski, Lin et al., 1997) discuss specific analysis steps to determine if hypothesized theoretical variables are causally related to (mediate) outcomes. These steps involve determining if a program or policy produced change on the hypothesized mediating factor; if this factor was related to outcome; and finally, to what extent outcomes were due to this factor vs. other issues.

Comprehensiveness

One of the central ways in which evaluation of programs intended for wide-scale implementation is different than other evaluations is in their need to be more comprehensive. In other words the scope of the evaluation needs to be broad, it can't just focus on narrow effects of a few variables.

This need arises from the complexity of programs which are ready for translation, the multi-level, contextual issues discussed above, and the importance of addressing concerns of multiple stakeholders and decision makers.

I will illustrate this point with a story. Imagine, as many scientists and citizens hope, that an amazingly effective new education program for HIV prevention is studied in a double-blind RCT efficacy study which demonstrated a large effect size – a 50% reduction in HIV rates compared to traditional HIV education.

This exciting breakthrough would then need to be put into practice to actually impact public health. Here is where the story gets interesting, and where the enormous impact of other behavioral, social, economic, and policy factors come into play. Further assume that the government and NGOs combine forces and resources in an unprecedented manner to rush this program into widespread use. Table 2 describes realistic to optimistic estimates of the actual impact of a nationwide dissemination effort to promote use of this breakthrough program. The right-hand column of Table 2 shows the bottom line public health impact or percent of all at-risk persons who would benefit from such an effort.

The left-hand column summarizes the series of steps involved in translating any basic science breakthrough into real-world practice. The second column displays the "success rate" for that step, and I have used estimates that vary from 40-60% for each stage to bracket the likely overall impact. For most steps, a 40-60% success rate would be considered a very good result from a nationwide campaign over a one- to two-year period; and especially if the 40-60% impacted were representative and included those most at risk (which unfortunately is often not the case).

TABLE 2

The Reality of Translating an Evidence-Based Educational Intervention into Practice

Translation Step	Success Rate	Population-Wide Impact
Educational Settings that Participate	40% - 60%	40-60%
Educators Who Adopt Program	40% - 60%	16% - 36%
Citizens/Students Who Participate	40% - 60%	6% - 22%
Participant Adherence	40% - 60%	2% - 13%
RCT EFFICACY RESULTS	**40% - 60%**	0.8% - 8%
Continued Longer-term Effects	40% - 60%	0.3% - 5%

If 40-60% of all educational settings (e.g., schools, community centers, and healthcare clinics) were to adopt this new treatment approach, that would be a phenomenal success. To accomplish this, a terribly convincing case would need to be made to diverse organizations that would include public, private, and religious schools; outpatient and hospital settings; community health centers, etc. – most of which have their own lengthy approval processes and many of which are under-resourced already.

The second row in Table 2 illustrates the impact of educator reactions to a newly approved

program, and again optimistically assumes that 40-60% of educators will adopt this new curriculum. The remaining rows of Table 2 illustrate the impact of later steps in this sequential story of the national roll-out of a new program. Only in the third and following rows of Table 2 do we even begin to include the impact of citizen/student reactions to such a program – which is, however, the only information in Table 2 that is reported in many published studies or in the media.

Three points should be made in summary: 1) The 40-60% estimates for the percent of citizens/students who would follow recommended practices; and who would continue to maintain benefits long term are likely to be over-estimates. 2) Only in the next to last row do the results of the groundbreaking initial study come into play – the issues in all the other rows are typically ignored in a basic trial designed to answer only the narrow question of whether a treatment will work under optimal conditions. 3) Finally, the "bottom line" impact after one to two years is that approximately 0.3-5% of the population would benefit in a lasting way from this revolutionary educational breakthrough.

Lessons Learned from Example (and Translational Research)

The purpose of this exercise is not to disparage educational approaches – the same issues apply to real-world application of behavioral, health care, or socio-economic interventions. The point is that evidence needs to expand beyond the narrow domain of studying only the impact on a single primary dependent variable. There is also an important, but optimistic, message embedded in Table 2.

This message is that there are numerous opportunities – represented by EACH row in Table 2 – to enhance the ultimate success rate in the bottom right of the table. Improving any of the steps shown could also substantially increase the overall benefit. These various steps also make apparent the opportunities for transdisciplinary collaboration to address translation issues – the potential contributions of diverse fields such as social marketing, health communication, behavioral approaches to adherence, risk and decision analysis, economics, and educational policy are apparent.

Another lesson learned, especially when conducting evaluations with limited budgets, is that it is costly and often burdensome to collect quantitative measures on all of the issues in Table 1. In addition, validated scales or practical instruments frequently do not exist for the specific evaluation questions in a particular project. In such cases, using a multi-method approach that includes qualitative assessment (Crabtree & Miller, 1999) can help provide a more complete evaluation. Use of qualitative, semi-structured interviews are particularly helpful in elucidating reasons for results and explaining "trouble spots" uncovered during evaluation (e.g., why certain subgroups do not choose to participate, why certain program components are not delivered consistently).

An excellent example of using quantitative and qualitative assessments together comes from the WISEWOMAN project to reduce cardiovascular risk among low-income women in the United States (Besculides, Zaveri et al., 2006). The investigators first used quantitative measures from the RE-AIM model (see below and also http://www.re-aim.org) to evaluate program reach, effectiveness, adoption, implementation, and maintenance similar to the steps in Table 2. From these measures, they identified sites that were especially high or low on the RE-AIM dimensions, and conducted observations, qualitative interviews, and focus groups within these sites to better understand factors associated with success.

Cost

Data on program costs and cost-effectiveness are essential to comprehensive evaluation, but one of the least frequently reported types of data in evaluations (Glasgow, Klesges et al., 2004). This is especially unfortunate since program cost is one of the first questions that decision and policy makers ask, and is often a major barrier to dissemination. Part of the reason there have not been more cost analyses is that program developers and researchers have felt overwhelmed by the complexity, magnitude of the task, and the time and costs involved in performing economic analyses. Fortunately, recent advances and simplifications are now available that do not attempt to answer every economic issue, but restrict focus to issues of the costs of program implementation (or to replicate the program in other settings) and/or the cost per unit change. Such models are practical for most evaluations (Ritzwoller, Toobert et al., 2006), answer the questions that decision makers usually have, and do not require a great deal of economist time (unlike more complicated issues such as determining cost-benefit).

Customization vs. Fidelity

One of the current program implementation and evaluation areas of active research and debate is how to resolve the inherent tension between customizing or adapting programs to local situations, using principles of participatory research (Viswanathan, Ammerman et al., 2004) to make policies/programs culturally relevant (Castro, Barrera, Jr. et al., 2004), and the need to maintain fidelity to an evidence-based program (Bellg, Borrelli et al., 2004) when translating research into practice. There is agreement that the extremes on either end of this continuum are not good. For example, having users make wholesale modifications to evidence-based interventions without sufficient justification or because they are not experienced with a certain component (e.g., omitting role playing from skills training); or on the other hand, expecting an under-resourced rural mental health clinic that serves a low-income, low-literacy population to conduct a program exactly as it was in a research study and to use precisely the same recruitment and educational materials would neither be expected to work.

Most promising approaches to balance customization and fidelity seem either to be driven by theoretical principles or a more pragmatic "essential components" evaluation analysis. The theoretical (vs. procedural) fidelity" approach (Rovniak, Hovell et al., 2005) evaluates program implementation based upon what theoretical principles are addressed by a given component. If a modification retains a similar emphasis on the theoretically important principle, then the approach is said to have theoretical fidelity. For example, a local community organization might choose to implement a "follow-up contact and support" principle by using case workers rather than public health nurses to conduct follow-up home visits.

The other approach is to have either experienced program developers or a panel of experts with practical experience in the content area and knowledge of the research-based program to designate a priori some intervention components as essential or necessary and other components to be modifiable (Ory, Mier et al., 2007).

Table 3 summarizes several key challenges faced when evaluating policies and programs in applied settings or assessing potential for dissemination. It uses the RE-AIM model (Glasgow & Linnan, 2007); http://www.re-aim.org) to consider both common challenges and possible solutions. RE-AIM is an acronym to help remind evaluators, planners, decision and policy makers, and those implementing programs about the various steps and levels necessary for successful programs and policies, as illustrated in Tables 1 and 2. It stands for Reach, (participation rate and representativeness), Effectiveness (intended and unintended outcomes), Adoption (settings and personnel that will deliver the program), Implementation (consistency and quality of program delivery), and Maintenance (long-term effects and sustainability).

TABLE 3
Common Challenges in Implementing and Evaluating Programs and Policies, and Possible Remedies

CHALLENGE	REMEDY
Reach: Not including a relevant, high risk, or representative sample or being able to evaluate representativeness	Use population-based recruitment or over-recruit high-risk subgroups Report on participation rate, exclusions and representativeness Avoid too many exclusion criteria
Effectiveness: Not thoroughly understanding outcomes or how they come about: - Assessing outcomes too narrowly - No knowledge of mediators - No assessment of moderator variables - Inadequate control conditions to rule out alternative hypotheses	Assess broad set of outcomes including possible negative ones Include measures of hypothesized mediators Conduct subgroup analyses to identify moderator effects Design control condition to fit your question
Adoption: Program only implemented in high functioning, optimal settings Program not ever adopted or endorsed – or only used in research or original development settings	Involve potential adoptee using CBPR* principles beginning with initial design phase Approach a representative or broad group of settings early on when revision is still possible – and report on setting exclusions, participation, and representativeness
Implementation: Program or policy not delivered as intended Not able to answer key questions about costs, time or staff requirements Deciding if a program adaptation or customization is good or bad.	Assess if too complicated, too intensive, or not compatible with other duties to be delivered consistently Systematically vary staff characteristics and evaluate staff impact as well as costs Specify *a priori* the critical theoretical components Identify essential elements that cannot be changed and those that can be adapted
Maintenance: Program or policy effects not maintained over time Substantial attrition of settings, delivery staff and/or participants over time	Include maintenance phase in both protocol and in evaluation plan Plan for institutionalization, sustainability, and dissemination and their evaluation Take steps to minimize attrition, address attrition using appropriate methods, evaluate and report impact of attrition

*CBPR = Community-based participatory research
See www.re-aim.org for more detailed discussion of RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) framework

The chief challenge to assessing "*reach*" is that too often evaluations include only participants who are easy to access, most likely to benefit, or especially motivated, and this results in setting unrealistically high recruitment expectations for translation. Another danger is of casting too narrow a net in evaluating results (*effectiveness*), focusing only on restricted outcomes and omitting measures of possible negative effects, mediating variables and process measures that can help understand why and how program/policy effects (or lack of effects) occur. Table 3 presents ways to broaden this perspective. Decision makers are concerned about impact on participants like those in their setting. Moderator analyses or evaluations of whether a program is differentially effective across subgroups that differ on important psychosocial and demographic factors can help to clarify applicability.

More evaluations should be conducted in representative or low resource settings. Evaluators should give equal priority to the recruitment and representativeness of settings (*adoption*) as is given to describing the representativeness of individual participants. Two crucial *implementation* issues often present challenges to interpretation and research translation. The first is failure to identify the characteristics of settings and staff that are able to successfully implement programs or policies. Staff characteristics that may moderate implementation include expertise, education, training, age, race/ethnicity, gender, experience, and similarity to the target audience. The second issue is that estimates of the program costs are often not available.

Resolving the tension between fidelity (delivering a program exactly as in a guideline or curriculum protocol) and customization or adaptation to local settings, culture and history are among the most important evaluation challenges. As discussed above, recommended approaches include specifying key or critical components of a program, and evaluating delivery of the theoretical principles that are hypothesized to lead to desired outcomes. Logic models (Glasgow & Linnan, 2007) are useful for depicting predicted relationships and in guiding measurement decisions.

There is a dearth of information on *maintenance* or sustainability of programs at the setting level. We need much greater understanding of the extent to which implementation is continued unchanged in the original setting, is adapted or discontinued (Goodman, McLeroy et al., 1993). At the individual level, participant attrition is a common challenge. I recommend analyzing the characteristics of those present (vs. non-participants) at follow-up assessments, and then deciding which imputation methods are most appropriate for that particular missing data situation.

The key to successfully overcoming the challenges summarized in Table 3 is to plan for, and anticipate trouble spots (Klesges, Estabrooks et al., 2005). The world is complex and program effects are often context-dependent. Our evaluations should reflect this complexity, and reports should transparently describe program challenges, adaptations, and contextual issues so that others can better learn from and judge whether a given policy or program will work in their situation.

This article has summarized key issues, challenges, and lessons learned in evaluating programs and policies intended for broader translation across a variety of content areas, applied in a wide range of settings, and for different populations. Trade-offs must be considered when planning formative, outcome, impact, and process evaluation efforts in a complex world (Linnan & Steckler, 2002; Glasgow & Linnan, 2007). Some of these interventions and policies will prove effective and should be considered for sustainability and translation; others will not. Data collected to answer the questions above will reveal program effects, limitations, processes and pathways of change; as well as insights about how to improve the theory guiding the policy or program. Such broad based evaluation approaches should help lead to innovations that are more effective, efficient, broadly applicable, and sustainable.

References

Baranowski, T., Lin, L. S., Wetter, D. W., Resnicow, K., & Davis, H. M. (1997). Theory as mediating variables: Why aren't community interventions working as desired? *Annals of Epidemiology, 7*, S89-S95.

Basch, C. E., Sliepcevich, E. M., & Gold, R. S. (1985). Avoiding Type III errors in health education program evaluations. *Health Education Quarterly, 12*, 315-331.

Bellg, A. J., Borrelli, B., Resnick, B., Ogedegbe, G., Hecht, J., Ernst, D., & et al (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the Behavior Change Consortium. *Health Psychology, 23*, 443-451.

Berwick, D. M. (1996). A primer on leading the improvement of systems. *BMJ, 312*, 619-622.

Besculides, M., Zaveri, H., Farris, R., & Will, J. (2006). Identifying best practices for WISEWOMAN programs using a mixed-methods evaluation. *Preventing Chronic Disease, 3*, 1-9.

Castro, F. G., Barrera, M., Jr., & Martinez, C. R., Jr. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science, 5*, 41-45.

Crabtree, B. F. & Miller, W. L. (1999). Depth interviewing. In B.F.Crabtree & W. L. Miller (Eds.), *Doing Qualitative Research* (pp. 89-107). Thousand Oaks, CA: Sage Publications, Inc.

Glasgow, R. E. & Emmons, K. M. (2007). How can we increase translation of research into practice? *Annual Review of Public Health, 28*, 413-433.

Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Bull, S. S., & Estabrooks, P. (2004). The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine, 27*, 3-12.

Glasgow, R. E. & Linnan, L. (2007). Evaluation of theory-based interventions. In K.Glanz (Ed.), *Health Education: Theory, research and Practice* (4th Edition ed., Hoboken, NJ: Jossey-Bass Publishers.

Goodman, R. M., McLeroy, K. R., Steckler, A., & Hoyle, R. (1993). Development of level of institutionalization scales for health promotion programs. *Health Education Quarterly, 20*, 161-178.

Green, L. W. & Ottosen, J. M. (2004). *From efficacy to effectiveness to community and back: Evidence-based practice vs. practice-based evidence* Proceedings from conference: From Clinical Trials to Community: The Science of Translating Diabetes and Obesity Research: National Institutes of Diabetes, Digestive and Kidney Diseases.

Institute of Medicine & Committee on Quality Health Care in America (2003). *Crossing the quality chasm: A new health system for the 21st century* Washington, DC: National Academies Press.

Kaplan, R. M. (2003). The significance of quality of life in health care. *Quality of Life Research, 12*, 3-16.

Kingdon, J. (1995). *Agendas, alternatives, and public policy.* (2nd ed.) New York: Harper Collins.

Klesges, L. M., Estabrooks, P. A., Glasgow, R. E., & Dzewaltowski, D. (2005). Beginning with the application in mind: Designing and planning health behavior change interventions to enhance dissemination. *Annals of Behavioral Medicine, 29*, 66S-75S.

Linnan, L. & Steckler, A. (2002). Process evaluation and public health interventions: An overview. In A.Steckler & L. Linnan (Eds.), *Process Evaluation in Public Health Interventions and Research* (pp. 1-23). San Francisco: Jossey-Bass Publishers.

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003). The quality of health care delivered to adults in the United States. *The New England Journal of Medicine, 348*, 2635-2645.

Orleans, C. T. (2000). Promoting the maintenance of health behavior change: Recommendations for the next generation of research and practice. *Health Psychology, 19*, 76-83.

Ory, M. G., Mier, N., Sharkey, J. R., & Anderson, L. A. (2007). Translating science into public health practice: Lessons from physical activity interventions. *Alzheimers and Dementia, 3*, S57.

Ritzwoller, D. P., Toobert, D., Sukhanova, A., & Glasgow, R. E. (2006). Economic analysis of the Mediterranean Lifestyle Program for Postmenopausal women with diabetes. *Diabetes Educator, 32*, 761-769.

Rovniak, L. S., Hovell, M. F., Wojcik, J. R., Winett, R. A., & Martinez-Donate, A. P. (2005). Enhancing theoretical fidelity: An e-mail-based walking program demonstration. *American Journal of Health Promotion, 20*, 85-95.

Viswanathan, M., Ammerman, A., Eng, E., Gartlehner, G., Lohr, K. N., Griffith, D., & et al (2004). *Community-Based Participatory Research: Assessing the Evidence. Evidence Report/Technology Assessment No. 99* (Rep. No. AHRQ Pub No. 04-E022-2). Rockville, MD: Agency for Healthcare Research and Qualilty.

The leading European evaluator, Franz Leeuw, draws attention to what he calls the "performance paradox". Goals are set, targets are set and measures are developed. Make people and agencies accountable for achieving measured targets and the goals are achieved. Except that in reality it doesn't quite pan out in this way. The performance paradox occurs when you hit the targets but fail to meet the goal. Targets are motivators and if people and agencies are made accountable for those targets then they become very good at hitting those targets. Inevitably the focus shifts from the goals to the targets and over time hitting the targets become the main objective, and the original goal falls by the wayside. Goal displacement occurs – the goals are displaced by the targets. The trouble is that the measures can never be as rich or as diverse as the goals. Nor can they reflect the important nuances and perspectives by which complex goals are achieved. Therefore focusing purely on targets and measures means that the goal is never fulfilled. Hence the performance paradox. Mariana's contribution discusses this paradox in relation to the Millennium Development Goals, assesses its possible impact ... and how to avoid it.

Monitoring and evaluation, knowledge management, and public policy – the x, y and z of the Millennium Development Goals equation

Mariana Stirbu

This article aims to open a theoretical discussion on the complex relationship between several seemingly unrelated concepts: Millennium Development Goals, on the one hand, and monitoring and evaluation, public policy, and knowledge management, on the other. It explores the ways in which a better understanding of this relationship can contribute to better MDG achievement, add value to development cooperation, and in particular, to achieving better development results.

The issues...

Many development organisations work today towards achieving the Millennium Development Goals (MDGs) – a set of time-bound, specific and quantifiable goals, containing 18 targets and 48 indicators to measure progress. Many have, at least partially, aligned their organisational goals to the MDGs and increasingly measure their degree of success in terms of their impact on one or more of them. To be successful in this endeavour, development organisations need to have more than just efficient internal coordination; they need an improved ability to be responsive to the situation of the poor and marginalised, carry out their work based on evidence, and use evidence to influence policy debates and processes (Hovland, 2003). Moving upstream into policy level work requires attention to policy design but also to issues of implementation and evaluation, which includes assessing the organisational impact on policy change and ultimately on social change. In this context, three elements seem to be key to ensuring MDG achievement:

1. monitoring and evaluation,
2. knowledge management, and
3. evidence-based policy.

The sections that follow address each of these key elements in turn and then look at how they relate to each other in relation to MDGs based on a case study of a UNICEF regional partnership program.

Monitoring and evaluation...of what?

Monitoring and evaluation in relation to MDGs includes at least two main aspects: 1) monitoring and evaluation of MDG goals and targets *per se*, and 2) monitoring and evaluation of policy change that contributes to MDG achievement.

Monitoring and evaluation of MDG goals and targets. There is a significant body of critique of the MDG conceptual and methodological approach, especially in relation to definitions, targets and indicators. Several key problems are pointed out, of which only some are summarised below (drawing almost exclusively on Saith, 2006).

Regarding the MDG of reducing poverty and hunger – one of the most important among the eight goals, limitations stem from the choice of the international one-dollar-a-day definition and measurement of poverty, with its implied focus on the inability to meet basic needs and a disregard for the multi-dimensional and relational nature of poverty (Elson, 2004). This approach, applied uniformly across the board, results in a situation where even if millions are alleviated from poverty, their situation will barely improve if living costs are higher than the international poverty line assumes. This limitation is acknowledged in the recently launched Millennium Development Goals Report 2007 (UN, 2007), which proposes a re-assessment of the international poverty line. The MDG of reducing poverty and hunger also has an intrinsic focus on poverty rather than poverty and inequality. Yet, inequality, in both income and wealth, has been increasing over the last several decades (Sundaram, 2007), and acting as a counter force to poverty reduction.

Another difficulty lies in the implied focus of MDGs on quantitative indicators and thus its disregard for the qualitative aspects of development. For instance, educational enrolment indicators create distortions by ignoring issues of quality of outcomes and of resources committed per child in the educational process, whereas the exclusive focus on primary education, with its myopic short-term perspective and disconnect from development processes, leads to a situation where the future of millions of children is being devalued. There is also no indicator to measure progress in reducing gendered stereotypes in the content of education (Elson, 2004).

There are also limitations in relation to gender equality and empowerment, which is measured by sex parity in primary education. It distracts attention from important issues of equality of sexual and reproductive rights relating to deep structural factors and of power relations, which are at the root of the perpetration of gender inequality and oppression.

There is also the generic concern over the usefulness of targeting given the different levels at which the countries start, as well as the potential for distorting behaviour or creating perverse incentives. For instance, there might be an incentive to allocate resources to those closest to the poverty line, rather than the poorest, in order to maximise the impact of a given budget on poverty reduction measured by the MDG poverty headcount ratio; or to switch expenditures from non-targeted items (e.g. the disabled, orphaned etc.) to other targeted items in order to be able to report the most favourable results for MDG outcomes. Related to this is the problem of baselines. Many countries start from very low 1990 levels, meaning that even if the target is achieved, the situation will not have improved substantially, pointing to a state of extreme complacency of the MDGs (Koehler, 2007). Last but not least, there is the problem of lack of data to ensure consistent monitoring and evaluation.

Compared to the critique of complacency, usefulness, and potential of the MDGs to effect change, the issue of monitoring and evaluating MDG targets and outcomes presents fewer challenges. This is because a list of indicators for monitoring and evaluation – meaningful at least in relation to the targets, if not the goals – does exist and has been sufficiently institutionalised.

Monitoring and evaluation of policy change. The monitoring and evaluation of policy change on the other hand is more challenging. Several issues are pointed out (drawing almost

exclusively on Mahmudov, 2007, and Gurthrie et al., 2005): *implementation* (in reference to complexity, external factors, timeframe, among others), *sustainability, attribution* and, again, *indicators*.

The path to policy change is complex and iterative, and often involves more than just getting legislators to change a law; it could sometimes involve efforts aimed at stopping a policy initiative if it is deemed detrimental to beneficiaries. Numerous external factors, such as the political and economic environment influence and shape policy implementation. Reforms are usually multi-faceted, which means that no particular effect or impact is easily attributable to a particular intervention as simultaneous shifts in the political environment, donor assistance etc. also have a bearing on reforms. The time frame for evaluation of policy change is also critical and it is different from that required for project work. It often cannot be assessed within a short period of time. Moreover, the adoption of a policy is only the first step towards addressing the achievement of an MDG; it does not in and by itself guarantee the establishment of the institutional, policy environment and, ultimately, the occurrence of social change.

Some development organisations have developed a series of indicators to measure policy reform; yet, they chronically lag behind in developing indicators for measuring organisational *impact on policy change*, due to the difficulty of establishing the cause-effect relationship arising from the multiplicity of policy players who do not necessarily share the same motivations. The contributions of different players create a problem of attribution, which ultimately hampers lesson-learning and further policy work. Therefore, Mahmudov, 2007, and Gurthrie et al., 2005 propose focusing on measuring an organisation's *contribution to the policy environment* rather than trying to distinguish impacts that can be directly attributed to a single organisation. Under this approach, assessing policy change implies flipping the question from: *What intended and unintended outcomes it has produced? to How did the organisation's work influence the policy environment? or How successful was the organisation in taking steps towards changing the policy?*

And knowledge management…for what?

A first step towards designing a set of meaningful indicators to help development organisations in assessing their development policy work and informing future policy advocacy would be to assess the processes of sharing of information with partners. This in turn is part of the overall efforts of building partnerships for development as a step counting towards influencing and changing a policy. In general, sharing of information, along with the sharing of resources and common objectives can influence and drive all phases of the policy cycle: planning, implementation and evaluation (Jackson and Kassam, 1998) through increasing the responsiveness of development partners to the situation of the disadvantaged, and developing partners' engagement in policy debates and policy processes and impact on policy (Hovland, 2003).

Sharing of information is an integral component of the processes of knowledge management, an emerging concept in organisation theory. Knowledge management is commonly defined as having two roles: 1) identifying, managing and sharing an organisation's knowledge and 2) enabling professionals to create knowledge collectively in order to achieve the organisational objectives (UN Staff College). Thus knowledge management enables organisations to engage better in policy debates and policy processes. This strategy has become increasingly important in the context of MDG achievement in the poorest performing regions, such as South Asia (see e.g. UNESCAP, 2006; UNICEF 2006a).

Good progress on the MDGs is especially relevant to the South Asia region, as more than half of the goals are directly linked to the situation of children, and children account for almost half of the total population in South Asia (UNICEF, 2006b). Although the MDG methodology provides ready-made targets, indicators and templates to monitor the situation, as noted above, there are many challenges to adequate monitoring and evaluation of MDG achievements.

This is particularly true for the MDG impacts of development cooperation programs in South Asia both for the governments and for the international community. This partially relates to the challenges of effectively sharing data and information on the situation of children and women (including from monitoring and evaluation), to support informed decision-making, policy analysis and advocacy around regional and national development. Moreover, the type of knowledge (data and information) and the particular knowledge management strategy adopted determines to a large extent an organisation's ability to influence policy.

Two types of knowledge can theoretically be distinguished as relevant in relation to MDGs: knowledge of the field (bottom-up learning) and knowledge of high-level negotiation processes (top-down learning) (Hovland, 2003). The first is important in boosting legitimacy and influencing policy. The second is important because for an organisation to have an impact on policy, it needs to know the channels through which to influence policy (ibid.). All of this relates principally to the first role of knowledge management – identifying, managing and sharing knowledge.

And evidence-based policy...for whom?

However, understanding MDG achievement in South Asia also requires going beyond the obvious targets and indicators, and overcoming the afore-mentioned MDG limitations by analysing and understanding the deep-seated factors determining development processes and challenges. This means using the data and information that exists to create new knowledge that would enable a development organisation to understand the bottlenecks in MDG achievement and thus be able to engage meaningfully in policy discourse and policy processes. This relates to the second role of knowledge management.

The development literature (see e.g. ADB, 2006; UN, 2006; WB 2005; UNICEF 2006; Birdsall, 2006) highlights at least three major development challenges – which are at the same time reasons behind the slow MDG achievement – in South Asia: human poverty and deprivation; inequality and disparity; and social exclusion. Some of these are not directly measured by the MDGs, especially the non-obvious dimensions of deprivation, inequality and exclusion. Thus, while the MDGs are straightforward and include targets and indicators to measure progress, they are not completely reflective of the development situation in the region. This is especially critical for categories of children whose situation improves – whether it is the poorest of the poor, the most marginalised, the most disadvantaged, the most neglected and abused, or just those close to the poverty line who are easier to target and lift above. Reducing poverty does not necessarily mean reducing deprivation, inequality, and social exclusion. The latter are deep-rooted phenomena in South Asia, which often hinder MDG achievement, as many studies have shown (e.g. Kabeer, 2006; Namala, 2006; Koehler and Keane, 2006). For instance, while the MDG on primary education records good progress, a look at intra-group variations in enrolment rates shows that very often people from the socially excluded groups have much lower rates of achievement than the others. Furthermore, within the same socially excluded group, some are more excluded than the others.

As one example among many, a recent study on social exclusion in Pakistan (OPM, 2005) shows that within the socially excluded group, gender gaps are wider than within the larger population. In particular, among sharecroppers, net enrolment rates for girls are much lower than those for boys than the enrolment rates between girls and boys in the total population. Looking at processes of exclusion is important for MDG achievement for the marginalised and disadvantaged groups, otherwise the risk of leaving them outside of mainstream development processes is high. These examples point to the need for development organisations to create new knowledge in order to challenge existing paradigms and approaches. In this context, examining the broader context, including the unintended outcomes of development cooperation programs, would help avoid common pitfalls and simplistic models where it

is assumed that inputs yield predictable outputs, and redirect attention more to the policy environment and policy change.

Joining up the three key elements...the value added?

A change in information needs of development partners will occur as they shift away from simple project approaches towards an emphasis on policy change, institutional reform and inter-agency partnerships (Moss et al., 1998, cited in Hovland, 2003). Given the non-linearity of policy processes, the capturing of intangible outcomes, such as policy, environment and institutional change requires drawing on rapid information feedback as well as new forms of field-level information (ibid.). These new forms of information, i.e. enhanced knowledge about development processes and outcomes in the field, could better equip researchers, policy-makers and partners to respond with adequate policy. This implies being open to learning and innovation at the same time.

Thus, equitable MDG achievement would require capitalising more on the second role of knowledge management – i.e. creating knowledge, including on the non-obvious dimensions of deprivation, exclusion and inequality, and sharing this knowledge with the development partners. This would ultimately increase the responsiveness of development cooperation to the situation of the disadvantaged and strengthen the engagement of development partners in informed policy debate and policy processes.

To sum it up, greater interaction between the three seemingly unrelated elements presented above – monitoring and evaluation, knowledge management and evidence-based policy – would ultimately provide for more disaggregated data and evidence on MDG achievement. It would help identify the policy channels and spaces through which to challenge the limitations of the MDGs. It would help advocate for the adoption of more ambitious and well-founded development policy goals and targets. How this triple-bottom approach has been used by UNICEF and what results it has generated is illustrated below.

Case study...UNICEF-AusAID Partnership Agreement and its intermediate results for South Asia

In its Medium Term Strategic Plan 2006-09 (MTSP), UNICEF added a new focus area of work – FA5, which proposes to "strengthen the capacities of States and societies to design and implement social and economic policies, legislative measures and budgetary allocations that enable them to meet their obligations...[as]...UNICEF is uniquely positioned to generate knowledge about the situation of children's and women's rights and to advocate and to promote partnership for their fulfilment" (UN, E/icef/2005/11:28-29). Under FA5, UNICEF's role is thus to become a more effective partner for advancing the MDGs. This is to be achieved by generating evidence and knowledge on the effects of deprivation on children and women, and by analysing the impact of public policies and stimulating dialogue on macro-level policies that guide national policy frameworks, legal reform and fiscal allocations affecting children and women. The MTSP proposes a series of indicators to measure progress under FA5, of which many focus on inputs/outputs. These indicators include number of countries with state-civil society organisations jointly managed budgets, number of countries generating thematic analyses on poverty, disparity etc.; number of countries carrying out reforms addressing the major gaps in the legislative, policy and institutional environment etc. The indicators relate to the organisation's *contribution to the policy environment* rather than to its *attributed impacts on policy change*.

To assist UNICEF in its efforts to advance the MDG agenda in South Asia, the Australian Government entered into a Partnership Agreement with UNICEF in August 2005. The Partnership Agreement emphasises the importance of two key components: analytical work and knowledge management. The analytical component is expected to contribute to a better

understanding of the factors hindering the timely delivery of the MDGs through analysis of causes, processes and effects of social exclusion and serve as basis for designing inclusive policy approaches. The knowledge management component is expected to identify and present solutions to MDG underachievement in order to get the lagging countries back on track. This includes analysis of existing policies, strategies and interventions that have addressed the delivery of basic social services to marginalised and excluded children. It also proposes capitalising on the linkages with monitoring and evaluation. Thus, under the Partnership Agreement, evidence-based public policy and knowledge management are assumed to be two mutually reinforcing elements of the efforts to accelerate MDG achievement. This theoretical model was proposed in the briefing note on knowledge management for UNICEF Regional Office for South Asia and subsequently refined in the UNICEF ROSA knowledge management report prepared by Sevenier (2007). After almost two years of implementation, the question is: *Does this assumption still hold, and what value added has this approach produced in terms of contributing to accelerating MDG achievement in South Asia?*

The Donor Progress and Utilisation Report 2006 to AusAID presents several key results achieved under the Partnership Agreement. These include

- increased understanding in UNICEF, governments and regional partners in South Asia of the underlying causes of MDG under-achievement in South Asia and of policy approaches to address social exclusion
- improved knowledge management in UNICEF for leveraging resources and policies for children's rights
- improved capacity in UNICEF for advocating with governments in South Asia for addressing the underlying causes of MDG under-achievement.

The analytical work by UNICEF Regional Office for South Asia (ROSA), which started with the assumption that social exclusion is the key underlying factor in slow MDG achievement, resulted in the unravelling of the multi-faceted-ness and multi-dimensionality of social exclusion, as well as produced the idea that a universalist approach to social service provision does not guarantee effective inclusion. Rather, special policy measures and interventions, to ensure that children and their families can actually claim and realise their rights to at least minimum basic services are required (Koehler, 2007). One example of such special measures are the educational grants for children of Dalits or other marginalised groups. These groups can be excluded from education, even if it is a constitutional right and provided free of charge, because of related costs. The educational grants cover costs to the family related to the child's education, such as uniform, transport and other expenses.

In the meantime, UNICEF ROSA has refined its approach and added the complex interplay of social exclusion with structural factors of inequality and disparity, fragility of states, and other factors to the list of causes of MDG under-achievement (Koehler, 2007). This is an illustration of how new knowledge is created to enable the organisation to better achieve its objectives.

The complexity of social exclusion and its interplay with the other factors has taken UNICEF ROSA in the direction of exploring how this complexity relates to individual MDG achievement but also to achievement across the MDGs. Understanding that "[s]haring is central to successful development cooperation" (Jackson and Kasssam, 1998), UNICEF ROSA organised an inter-agency MDGI consultation on exploring policy interfaces and synergies of outcomes in MDG-based planning for development in South Asia with UNFPA Country Support Team in Kathmandu and UNDP Regional Centre in Colombo. The discussions produced additional insights on the causes behind the slow individual MDG achievement and resulted in an informed exchange around macro and sectoral policies between the different participating agencies. The consultations arrived at several significant conclusions, summarised below (drawing almost exclusively on Koehler and Stirbu, 2007).

While being necessary goals, the MDGs are not *sufficient* to achieve socially equitable, inclusive and just development. Pursuing each goal in isolation of the others, without

understanding and exploring the linkages between them, and the synergies in outcomes that they can generate, risks overlooking crucial opportunities for results. Identifying and building on visible and invisible policy linkages can help un-block and strengthen the potential of the MDGs. It can also be used to address overarching factors, such as legal and regulatory frameworks, protection and security, on which MDG achievement ultimately depends. For instance, there are multiple synergies and benefits from linking education, health and nutrition or education and water and sanitation or decent work, gender equality and poverty reduction. At the same time, there are multiple benefits from exploring the linkages between the MDGs and the non-quantifiable commitments under the Millennium Declaration. For instance, child protection – although not itself a target – largely determines the progress of the educational goal especially in terms of the gender equity and equality aspects.

The consultations have at the same time brought out some differences in positions and approaches across the different agencies, such as financing the MDGs. These debates were carried out by means of an "open-argumentation dialogue" defined as a constructive dialogue among policy actors which offers "the possibility of reconciling or even superseding individual, branch or ideological contradictions" (Van der Knapp, 2004:20). The consultations resulted in "shared meaning" (Stone, 2002) or "shared notion" (Van der Knapp, 2004) between the agencies on policy theory with regard to the MDGs, i.e. on its key values, norms, and assumptions regarding causal linkages. Overall, the inter-agency work produced a conceptual shift through new and additional angles for policy analysis. For instance, the complexity of MDG achievement can be potentially unravelled by looking at MDGs holistically and also in relation to the commitments under the Millennium Declaration – a broader development framework than the MDGs. These have important implications for the monitoring and evaluation of the MDGs, because it challenges many of the traditional cause-effect assumptions. These findings and policy ideas have subsequently been used to influence decision-making in South Asia. At regional and national symposia and consultations a space was created for re-examining this new evidence and policy ideas to enable better-informed decisions on policy proposals around the MDGs. The re-examination can also influence the design and implementation of government programs, without necessarily diverting the government's focus on policy objectives and scope. Indeed it can relax what Chen (1990) calls, the rigidity constraints of goal-oriented policy models, by pre-emptying a search for the plausible unintended outcomes. So the policy discussions can contribute to re-entering and re-opening the "apparently closed policy systems" (Van der Knapp, 2004) shaped around the MDGs. This is an illustration of how knowledge creation and sharing could contribute to revisiting policy theory, including for purposes of improving future evaluation approaches to MDGs by favouring a move towards evaluation for innovation and learning.

In parallel to this conceptual and policy level work, UNICEF ROSA undertook a comprehensive survey of knowledge management practices (Sevenier, 2007) over the period November 2006 and April 2007. This resulted in a comprehensive review of current knowledge management practices in UNICEF in South Asia. The survey used an online questionnaire, complemented with individual open-ended interviews. Although the survey results are not generalisable due to several inherent limitations [1] it generated several useful findings in relation to the central research question posed by this article:

- One third of the respondents considered themselves familiar with program evaluations.
- Over half said that they used the findings of evaluations in their work.
- Over half said that they used lessons learned and good practices in their work, with two thirds accessing the lessons learned and good practices through personal networks and over half through annual reports.
- The majority said that they would like to have access to more lessons learned and good practices.
- Over one third of respondents considered that the most important knowledge

management-related function of their work was sharing and dissemination.
- Almost two thirds responded that they belonged to a community of practice – which are known to be platforms and media for the sharing of knowledge and information.

These responses point to the fact that UNICEF staff understand the linkages between knowledge management and evaluation. They also suggest that knowledge management is relevant to the evaluation function through the sharing and dissemination of evaluation findings, lessons learned and good practices and better utilisation of evaluations. The survey highlighted the need for linking up monitoring and evaluation with knowledge sharing for achieving better programmatic results.

Lessons learned… for evaluation theory and development cooperation in South Asia

As most of the work under the Partnership Agreement is still underway and as knowledge management is an emerging concept in organisational theory, it is too early to come to any firm conclusion. Moreover, no explicit evaluation has been undertaken to confirm the gains claimed above of joining up knowledge management, monitoring and evaluation and evidence-based policy. Yet, several conditions and propositions can be identified.

First, there is a strong linkage between public policy and knowledge management, as evidenced by the fact that the sharing of the knowledge accumulated by the different agencies working towards MDG achievement generated new insight about the inter-play between the MDGs, and a clearer understanding of how policy coherence can generate synergy of outcomes and accelerate MDG achievement.

Second, there is a strong linkage between knowledge management and monitoring and evaluation, as evidenced by the results of the knowledge management survey. Although the survey did not explore the ways in which these linkages could be reinforced, the responses suggest that one of the conditions is to equip the staff with the means to engage in sharing knowledge, as well as mainstream a culture of knowledge sharing into the core business practices.

Third, there are strong linkages between public policy and monitoring and evaluation, where the evidence gleaned from monitoring and evaluation provides the evidence needed for designing policy.. Thus, UNICEF ROSA's work on MDGs can be said to have indirectly contributed to the body of knowledge and evidence on theory-based evaluation. The collection of assumptions on which policy is based usually reflects a simplified reality, causing a tunnel vision, rigidity and fear of innovation, which pose an obstacle to understanding and improving policy (Van der Knapp, 2004). UNICEF ROSA's work generated new evidence on the need to depart from a strictly policy theory-based approach to evaluation of MDG-related policies and instead to focus more on 'evaluation for improvement' and 'evaluation for innovation'. While the MDGs clearly present a special case, these conclusions can be extrapolated to any development cooperation program aimed at influencing policy.

Several conditions seem necessary for this relationship between knowledge management, public policy and monitoring and evaluation to contribute to better MDG achievement. One of these is for organisations to be able to engage in 'open-argumentation dialogue', such as was achieved at the inter-agency MDGI consultations. Another is to engage in policy negotiation processes, including but not limited to evidence and knowledge needed to support one's approach.

Clearly, a lesson for policy-makers and development cooperation agencies is that evaluation of MDG performance in South Asia – a region characterised by complex issues of exclusion,

---

1   Such as the self-selection of respondents (whose participation was most probably driven by their genuine interest in knowledge management issues and thus the resulting bias), the small sample of the surveyed population, and the lack of a proportional sample based on various characteristics.

inequality and deprivation – need to embrace a 'learning-and-innovation' approach. Indeed the traditional approach might even debilitate the process of MDG achievement. The MDG-related issues are more complex than 'meets the eye' of the policy-maker or evaluator, which means that both policy theory and evaluation need to be open to innovation and learning.

References

ADB. (2007). *Asian Development Outlook 2007: Change amid Growth*, Asian Development Bank publication

Birdsall, Nancy. (2006). "The World is not flat: Inequality and injustice in our global economy". WIDER Annual Lecture 9. UNU-WIDER 2006

Chen, Huey-Tsyh. (1990). *Theory-driven evaluation,* Sage Publications, Newbury Park/London/New Delhi

Elson, Diane. (2004). The Millennium Development Goals: A feminist development economics perspective. 52nd Dies Natalis Address. The Institute of Social Studies.

Gurthrie, Kendall et al. (2005). "The Challenges of Assessing Advocacy: Strategies for a Prospective Approach to Evaluating Policy Change and Advocacy". October 2005. Funded for and prepared for The California Endowment

Hovland, Ingie. (2003). "Knowledge Management and Organisational Learning: An International Development Perspective", Overseas Development Institute, Working Paper 224

Jackson, Edward T. and Yusuf Kassam. (1998). Knowledge Shared; Participatory Evaluation in Development Cooperation. Kumariam Press. West Hartford. Connecticut

Kabeer, Naila. (2006). "Challenges to Social Policy and Social Protection of Excluded Children" In: *Social Policy: Towards Universal Coverage and Transformation for Achieving the MDGs* (CD). Kathmandu. May 2006

Koehler, Gabriele and Jennifer Keane. (2006). Analytical Report "Social Policy in South Asia: Towards Universal Coverage and Transformation for Achieving the MDGs" In: *Social Policy: Towards Universal Coverage and Transformation for Achieving the MDGs* (CD). Kathmandu. May 2006

Koehler, Gabriele and Mariana Stirbu. (2007). "MDG-based planning for development: policy interface and synergies of outcomes for South Asia." Report of MDG Policy and Program Review. 2007. UNICEF Regional Office for South Asia. Kathmandu

Koehler, Gabriele. (2007). "Social inclusion and the case for proactive and inclusive social policy in South Asia." Concept note for discussion and comments. UNICEF Regional Office for South Asia. February 2007

Koehler, Gabriele. (2007). "Social policy and social inclusion: special interventions in the political, economic and social domains", presentation at the South Asia Policy-Maker Symposium, UNICEF Regional Office for South Asia. 18-19 July 2007. Kathmandu

Mahmudov, Siraj. (2007). "Monitoring and Evaluation Policy Advocacy in UNICEF" In: *Internal Bulletin Something to Think About* in No. 43 sponsored by the Global Policy Section. Division of Policy and Planning, UNICEF New York

Namala, Annie. (2006). "Children and Cast-Based Discrimination" In: *Social Policy: Towards Universal Coverage and Transformation for Achieving the MDGs* (CD). Kathmandu. May 2006

OPM. (2005). "A Quantitative Assessment of Social Exclusion in Pakistan". Oxford Policy Management Institute. June 2005

Saith, Ashwani. (2006). "From Universal Values to Millennium Development Goals: Lost in Translation" in *Development and Change* 37(6): 1167-1199

Sevenier, Gaelle. (2007). "Learning form KM Experiences. Cases studies on KM initiatives in UNICEF South Asia, UN Regional Offices and Selected Agencies". Report of Information and Knowledge Management. UNICEF ROSA. April 2007. Kathmandu

Stone, D. (2002). Policy Paradox: *The Art of Political Decision Making*, New York: Norton

Sundaram, Jomo Kwame. (2007). "Global Income Disparities and the Need for Redistributive Economic and Social Policies in Developing Countries", presentation at the Regional Policy Maker Symposium organised by UNICEF ROSA, 18-19 July 2007, Kathmandu

UN. (2007). The Millennium Development Goals Report 2007. New York

UN. (2007). *World Development Situation and Prospects 2007*. United Nations publication

UN. Economic and Social Council. United Nations Children's Fund. E/icef/2005/11. "The UNICEF medium-term strategic plan, 2006-2009. Investing in children: The UNICEF contribution to poverty reduction and the Millennium Summit agenda."

UNESCAP. (2005). *A Future Within Reach: Reshaping institutions in a region of disparities to meet the Millennium Development Goals in Asia and the Pacific.* United Nations publication

UNICEF. (2006a). *The Millennium Development Goals: Progress and Challenges in South Asia.* UNICEF Regional Office for South Asia. Kathmandu

UNICEF. (2006b). *South Asia Booklet on Child-Relevant Socio-Economic Indicators*, UNICEF Regional Office for South Asia. Kathmandu.

United Nations System Staff College, http://www.unssc.org/web1

Van der Knapp, Peter. (2004). Theory-based evaluation and learning: Possibilities and challenges. *Evaluation*. Vol. 10(1): 16-34

World Bank. (2005). World Development Report 2006: *Equity and Development*. The International Development Bank for Reconstruction and Development/The World Bank and Oxford University Press