Evaluation
SOUTH ASIA

EDITED BY BOB WILLIAMS & MEENAKSHI SANKAR
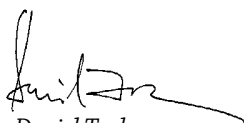
unicef

Evaluation

SOUTH ASIA

FOREWORD

Evaluation is an essential tool for making development programmes successful. Our ultimate measure of success is healthy and educated children, women surviving child birth, access to safe drinking water and sanitation and children growing up in a safe and protective environment. Evaluation is thus indispensable in assessing whether cooperation programmes actually contribute to improving results for children.

In the spirit of General Assembly and UN Economic and Social Council resolutions and as a member of the United Nations Evaluation Group, UNICEF is committed to the professionalisation of the evaluation function within the UN system and to developing national capacities for evaluation. This publication is our contribution to this challenging task and is intended to assist governments and practitioners in South Asia in strengthening evaluation and raising its importance as a core element in improving development cooperation results and achieving the Millennium Development Goals.

This publication gives impetus to developing evaluation capacities and supporting a culture of evaluation in South Asia. In bringing together various strands of thought in evaluation theory and practice from South Asia and elsewhere, it presents varied experiences and contributes to building institutional and individual capacity in evaluation.

A strong capacity and culture of evaluation is essential for achieving the Millennium Development Goals. With its targets prioritising primary education, maternal health, child survival and development, access to water and sanitation and more, the Millennium Development Goals are essentially about improving the situation of almost half the population of South Asia – its children. If this publication stimulates professional dialogue and action for the Millennium Development Goals, then our purpose is achieved.

Daniel Toole
Regional Director
UNICEF Regional Office for South Asia

CONTENTS

EDITORIAL – EVALUATION MATTERS

One of the great pleasures of editing a publication is to see the patterns emerging as the submitted articles develop through their various drafts and reviewers make their comments. It was great to see how seriously the authors sought to address all four aims of the publication:
- raise professional interest in evaluation in South Asia
- promote an understanding among public institutions and civil societies that evaluation feeds into knowledge management
- show evaluation as a strong tool for improving effectiveness, efficiency and sustainability of programmes
- demonstrate innovative evaluation methods.

These aims were broader than most evaluation publications, and consequently placed greater demand on authors.

We encouraged the authors to reflect on their practice and ideas and model the principles of evaluation as they shaped and reshaped their papers. We are delighted to see how much the authors rose to that challenge. We believe that these conversations between the editors and authors provided additional insights and lifted the overall quality of the debate around the critical issues discussed in each of the papers. Consequently, we decided to make these conversations 'public' as they offer a richer description of the contexts for readers of this book. We think this will help make this publication quite unique and have much broader appeal.

This is a publication that project managers, donors, policy makers, government agencies and members of broader civil society will find stimulating and useful. Collectively, the articles raise some important challenges to the direction in which evaluation in the developing world is heading, and helpfully provide some answers to those challenges.

It is very clear from reading the articles that the Millennium Development Goals, the ideas of Results Based Management, notions of evidence based practice, the increasing demands that evaluations "prove" impact and technocratic images of "knowledge management" are casting long and not always helpful shadows for evaluation practice in this region.

The articles – some explicitly and others implicitly – highlight key issues facing evaluation in this region. The needs of development agencies and the needs of those to whom a program is directed do not always coincide. Can these differences be accommodated, and if not whose interests should be paramount?

Increasingly, program managers are called upon by program funders to demonstrate the impact of their programs. This emphasis on 'proving' can undermine and sometimes even distract from program delivery efforts, where a focus on 'improving' could be more meaningful. It is not easy to design evaluations that both 'prove' and 'improve'. Also can an overemphasis on impact limit options for innovation and gradual achievement of sustainable results?

Russell Glasgow brings his long experience of transferring programs from one context to another and provides some valuable insights on how to avoid the traps for the unwary. But his article poses, for us, a big question. If "experimental designs" are difficult to generalize outside their own context, then the notion of randomized experimental design in particular as a universal "gold standard" for all settings may be problematic because the results might tell us little about the contexts, conditions and real world contingencies under which programs are likely to work. Therefore, encouraging use of particular approaches (e.g. randomly assigned experimental designs) for assessing performance of specific interventions may serve only a limited purpose. And if such assessments come at the end of the interventions, there is little value as the assessments do not feed into program design and delivery. Context is important

and there is a strong need for evaluations that assess not only ask what works, but also ask for whom in which circumstances.

Many of the methods described in this publication are clearly focused on 'improvement' yet will they survive this new pressure to 'prove'? And if so, what will be the consequences?

Another major trap is goal displacement, where the focus on specified, easily measurable "results" without any understanding of context distracts attention from what really needs to be done but is less easily measured. For instance, can we achieve the Millennium Development targets yet miss the goals? Mariana Stirbu drills deep for answers in her article and comes to some important conclusions that if taken up will certainly avoid the inevitable disappointments.

Right across the world, countries are struggling to come to grips with ideas of "evidence-based" policy and practice and "knowledge management". In just about every country the evaluation community scratches its collective head about the poor take up of evaluation results and the apparently arbitrary nature of the policy making process. At the same time, governments in South Asia are coming under increasing pressure to make better-informed decisions by putting the best available evidence at the centre of the policy process. Yet, it is rare to find successes in this area even in European and North American countries where resource availability is beyond what South Asian countries can ever hope for. Is this realistic? Is this fair? Contributions from Karma Galleg and Alex Nartea from Bhutan and Khadija Khan from Pakistan discuss some of the benefits and challenges in promoting a shift towards using evidence to improve quality of decision making in their country contexts.

This leads to another theme that we spotted - the balancing act between "bottom up" and "top down". Throughout the world, for the past couple of decades, the major trend in evaluation has been towards methods that enable improvement and learning. Learning within what is in reality often unknowable redefines what we understand as validity and accuracy. Sometimes good enough is good enough; fit for purpose and timely may be superior – and certainly more useful - than the ideal that arrives a year after the key decisions have been made.

Based usually on community development traditions or learning theories "bottom up" methods are evolutionary, varied, participatory, and highly context dependent – as exemplified in the articles by Ravi Jayakaran, Ana Coghlan and her colleagues. No one size fits all. Yet there seems to be strong preferences among development partners for particular methods to monitoring program performance. In between are Governments that try to manage the tension as best they can. The role of national evaluation organisations as a catalyst and a guiding hand in this situation is, as shown in the articles by Khadija Khan from Pakistan and Indra Tudawe and Mallika Samranayake from Sri Lanka. George Grob provides precious and invaluable guidance on how to manage these tensions by describing in clear terms what – on the basis of a lifetime of thought and experience - constitutes "good" evaluation, irrespective of time, situation and context.

These are all big themes. They would do justice to any evaluation publication: indeed any publication with an interest in using "evidence" to "improve" policies, programs and projects. So it is particularly exciting that these important issues are being addressed in the first South Asia focused evaluation publication.

This edition contains eight articles. Although they can be read in any order we have grouped them into three sections. The first section is all about overview – it describes the current context for evaluation and policy development and what can be learned from the wider world of evaluation and policy development. This section comprises papers by George Grob, Russell Glasgow and Marana Stirbu. The second section covers responses at a national level either by governments or civil society -Karma Galleg and Alex Nartea, Khadija Khan, Indra Tudawe and

Mallika Samranayake. The third section is about methods – Ravi Jayakaran and Ana Coghlan and her colleagues.

Finally our thanks to Sanjeev Sridharan, our reviewer, for helping the authors make their ideas even clearer, Rose Miller, our designer, for making the ideas easier to read, and to Soma de Silva, our guiding spirit, for providing the means by which the experiences and ideas can see the light of day.

Meenakshi Sankar
Bob Williams
Editors
September 2007

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-empting 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

PART TWO – THE OTHER PICTURE

In the ongoing debates about evaluation designs, especially those surrounding random controlled trials, one thing is often forgotten. No matter what the design, bad data is bad data. So the question should not be whether Method A is inherently better than Method B, but which approach will collect better data in the particular situation with the available resources. Two things spin off from this statement. Firstly people will give you good data if they trust you, feel part of the process and have a stake in the results being as accurate and valid as possible. Participatory methods of evaluation design are a potentially important way of doing this. Secondly all methods of data collection have their inherent biases. The more methods you use, the more the biases can cancel each other out. Hence the value of mixed methods. This paper is an excellent example of an evaluation that really understood these two principles.

Participatory and Mixed-Method Evaluation of MSM HIV/AIDS Programs in Bangladesh, Nepal and Indonesia

Anne T. Coghlan
Philippe Girault
Dimitri Prybylski

Introduction

Males who have sex with males (MSM)[1] in Asia often are at risk for HIV infection because of high levels of stigmatization and discrimination, low access to prevention, care and treatment services, low levels of condom use, and high turnover of sexual partners (MAP 2005). Family Health International (FHI), a U.S. non-profit organization that works primarily in the areas of HIV prevention and treatment, supports numerous MSM and MSM-related community-based organizations, government organizations, non-governmental organizations (NGOs), and regional programs throughout Asia. In 2005-2006, FHI commissioned the evaluation of its MSM programs in Bangladesh, Nepal and Indonesia. This paper describes some of the challenges faced in conducting MSM program evaluations in general, and the particular ways in which these challenges were addressed in the FHI evaluations. Specifically, it describes MSM and HIV vulnerability in Asia, the need to conduct and evaluate MSM programs, and some of the challenges in carrying out such evaluations. It then identifies the primary features of the three MSM program evaluations, including their purposes, objectives, major questions and methods. It then explores two methodological features in particular, their participatory approach and mixed-method design, and why and how these features were incorporated, and the benefits of doing so. Lastly, this paper invites donor agencies, local and international NGOs, and evaluators to consider integrating these features when conducting evaluations of not only MSM HIV/AIDS programs but different types of programs involving hard-to-reach populations in Asia.

---

1   In this paper, the term "MSM" refers to biological males who engage in sex with other biological males, regardless of the
     motivation for their sexual behavior, their age, and their sexual self-identification.

MSM and HIV Vulnerability in Asia

Same-sex sexual relations take place in all societies. In different surveys conducted among males from general populations in a number of Asian countries the prevalence of male-to-male-sex varies from 3.4% to 9.1%. (Beyrer 1995; Go 2004; Sittitrai 1994; van Griensven 2004).

The first cases of HIV and AIDS in several Asian countries were identified among MSM. In the few countries where studies have been conducted, HIV prevalence among MSM is typically elevated relative to the general population (Centers for Disease Control and Prevention, 2006; Choi 2003; Girault 2004; Go 2004; Pisani 2004; van Griensven 2005). Diversified commercial and noncommercial sexual networks, high turn-over of sexual partners, and high rates of unprotected anal intercourse increase MSM's risk of being infected with HIV and other sexually transmitted infections (STIs) (Choi 2003; Colby 2003; Girault 2004; van Griensven 2005). Stigmatization and social discrimination, with frequent physical and verbal violence, also increase the vulnerability of MSM (Preston 2004; Wilson 2004), particularly in areas where access to appropriate information, products such as condoms and water-based lubricant, STI services, counseling and testing (CT) services, and care, support and treatment services remains limited or nonexistent.

Many countries in Asia do not include MSM in their national surveillance systems. The information gap created by this omission perpetuates the vicious cycle of no data equals no problem, no problem equals no intervention, and no intervention equals no need to collect data (MAP 2005). Because of the paucity of data and innate discrimination, MSM are absent from the majority of Asian countries' HIV policies and program agendas (CDC-GAP/USAID-RDMA-FHI/APD 2005; Parker 1998). The poor attention to MSM in HIV/AIDS interventions has led not only to inadequate programs but also misperceptions, such as vaginal sex with females is risky behavior and male-to-male anal sex is not (MAP 2005).

As mentioned earlier, there are few HIV/AIDS programs in Asia targeting MSM. FHI supports existing local organizations, or "implementing agencies" (IAs), throughout Asia to deliver interventions aimed at reducing the number of new HIV infections among MSM and those who are transgendered. Although programs vary by country, general program strategies include increasing MSM access to knowledge, condoms, lubricants, and STI treatment services, and supporting positive environments that enable behavior change.

Opportunities and Challenges in Evaluating MSM Programs in Asia

As with many human service programs, there are several major reasons to evaluate MSM HIV/AIDS programs. These include:
- The need to improve program design and/or implementation
- The need to report and be accountable to funding agencies and other stakeholders
- The need to gain more knowledge about what works and to advocate for such programs.

There is an urgent need to evaluate MSM HIV/AIDS programs in Asia because these programs are relatively new and little is known about them. For example, Naz Foundation International has developed an MSM sexual health intervention model for South Asia (Naz Foundation International 2003), but this model has not yet been fully evaluated. In contrast, several evaluations have been conducted of MSM intervention models in Western countries (Johnson 2002; Wulfert 1996). But, there is little research on how these models can be adapted to different social and cultural environments in resource-poor countries in Asia. Thus, serious questions remain about the relevance and effectiveness of such models in Asian contexts.

Despite the numerous opportunities and reasons for evaluating MSM programs in Asia, there are several challenges in doing so. Although many human service programs face these challenges, they often are particularly common in MSM HIV/AIDS programs.

A Diverse Hard-to-Reach Population

Because sex between males is illegal, stigmatized and suppressed in many Asian countries, MSM social networks are frequently 'underground' or hidden, making MSM an often hard-to-reach population. This, in turn, makes MSM HIV/AIDS interventions particularly challenging to implement and evaluate. In addition, "males who have sex with males" encompasses diverse and fluid behaviors and social networks and, consequently, often it is difficult to identify individual MSM. Another challenge is to get MSM to openly and truthfully discuss their sexual orientation, behaviors and practices. Thus, it is particularly important when evaluating MSM programs that participants or respondents support the evaluation effort and are comfortable with data collection methods and the ways the resulting information will be used.

Diverse Stakeholder Needs

MSM programs often have multiple layers of stakeholders, many of whom have different information and evaluation needs. For example, FHI-supported MSM programs in Asia have the following primary stakeholders: the United States Agency for International Development (USAID) as the major funding agency, the FHI head office in the United States, the FHI regional office in Bangkok, the FHI in-country office, and, equally important, the local NGO or implementing agency, which includes both head office and field staff, program participants and other community members. In general, USAID and the FHI head and regional offices need evaluations of their major programs that investigate program effectiveness or the extent to which expected outcomes have been achieved. The FHI in-country office staff and the multiple stakeholders of the implementing agencies also are interested in process and implementation information (in addition to effectiveness), so as to improve programming activities. Meeting these diverse needs poses serious challenges, as the amount of human resources and time required to complete these types of comprehensive evaluations often are not available. Other challenges include the need to conduct these evaluations in ways that are most useful to the different stakeholders and presenting the voluminous evaluation results in ways that are relevant and meaningful to multiple audiences.

Attribution of Program Effects

As mentioned earlier, some audiences are interested in evaluations that determine program effectiveness and investigate issues relating to attribution. Some evaluation theorists argue that the question of attribution is best addressed through the use of experimental or quasi-experimental study designs. However, for many MSM HIV program contexts in Asia, the use of these designs is not feasible or desirable. Such methodologies often are too expensive and time consuming, and could divert limited resources from program implementation. Moreover, there are serious ethical issues in randomly assigning individuals or communities to an experimental group (individuals or communities that receive the intervention) and to a control group (individuals or communities that do not receive the intervention). Thus, HIV/AIDS program evaluation often is conducted using population-based data from surveillance or program-specific cross-sectional surveys with no external control groups. Respondents exposed to interventions then are compared against internal controls (those unexposed to programs) to help assess program effects. Nevertheless, assessing and substantiating the linkages between program activities and outcomes and impact remain a serious challenge, particularly as donor agencies are increasingly requiring that programs "prove" their effectiveness to justify their financial investments.

Description of the FHI Supported MSM Programs

A key purpose of the FHI evaluations was to identify program effects. Consequently, a decision was made to conduct the evaluations among the three longest running FHI-supported MSM programs in Asia, as this would allow adequate time for program effects to have taken place.

The three FHI-supported MSM programs were the Bandhu Social Welfare Society (BSWS) in Bangladesh, the Blue Diamond Society (BDS) in Nepal, and the Aski Stop AIDS (ASA) program in Indonesia. Within the ASA program, the IAs included in the evaluation were Persatuan Waria Kotamadya Surabaya (PERWAKOS), Yayasan Gaya Nusantara (GN), Yayasan Pelangi Kasih Nusantara (YPKN), and Yayasan Srikandi Sejati (YSS).

This purposive sample was appropriate for summarizing the major accomplishments, challenges, and lessons learned of FHI's most mature MSM programs in Asia. Although each organization was unique in its structure and operational context, all three FHI-supported components aimed to:

- Increase and sustain safer sex behavior and appropriate health seeking behavior through behavior change communication;
- Increase access to commodities such as condoms and lubricant and to services such STI treatment, counseling and testing, and HIV care, support and treatment;
- Create a favorable environment to support behavior change; and
- Strengthen the capacity of the implementing agencies' staff to manage and carry out the interventions.

Program activities conducted by each of the agencies included:

- Outreach and peer education;
- Establishment of drop-in centers or safe places where MSM socialize and also receive educational and health services;
- Condom and lubricant promotion and distribution;
- Community mobilization events;
- Referral services to STI treatment, counseling and testing, care and support, all specially sensitized and tailored to MSM; and
- Local and national-level advocacy.

A major difference between the IAs was their size and their operational structure. For instance, BSWS had nine offices in six cities throughout Bangladesh, BDS operated primarily in Kathmandu with one office and two drop-in centers, and the four IAs in Indonesia had one office each, covering a portion of the cities in which they were located. Another major difference was the amount of emphasis given to conducting advocacy activities. While BDS saw advocacy as a major function, some of the other organizations conducted advocacy as needed but in more moderate ways.

Major Features of the MSM Program Evaluations

Although each implementing organization was unique, the FHI-supported components were similar enough to allow the development of a common evaluation framework that consisted of the same general purposes, objectives, evaluation questions and data collection methods. However, organization-specific issues, concerns and questions were added to the common framework where needed.

Evaluation Purpose
The purpose of the evaluation was clarified through detailed discussions with FHI and IA staff and evaluation planning workshops conducted with program staff and participants. As defined in the workshops, the multiple purposes of the evaluations were to:

- Improve current program design, management, and implementation;

- Determine the overall effectiveness of the MSM program approach and its various components in achieving expected outcomes;
- Inform possible program expansion and determine ways to expand program coverage;
- Increase awareness of MSM issues and gain support from MSM communities, local and international NGOs, international development partners, and government agencies; and
- Inform future agreements between FHI and the IAs and the allocation of financial and technical assistance.

Evaluation Objectives

The overarching objectives for each of the three country evaluations were:

- To devise a logic model depicting the program's inputs, outputs and intermediate and long-term expected outcomes with input from multiple stakeholders.
- To conduct a process evaluation, determining the number and quality of services implemented and program coverage.
- To conduct an outcome evaluation, determining the extent to which expected outcomes have been achieved.
- To the extent possible, foster the evaluation's "process use" (Patton 1997) by enhancing communication and a shared understanding of the program among stakeholders, increasing engagement and ownership of the program by multiple stakeholders, nurturing an evaluation culture within the organizations, and developing internal evaluation capacity.

Major Evaluation Questions

The general questions answered in all three country evaluations included:

1. **Process Evaluation**
- What is the logic or "theory of change" underpinning the intervention?
- What activities were implemented and were they implemented as planned? Why or why not?
- Who has the intervention reached and what is the extent of program coverage?
- What is the level of quality of the services and commodities provided?
- What is the level of satisfaction by beneficiaries and what are their suggestions for improvement?
- What are the strengths and weaknesses of the individual interventions and the overall program?

2. **Outcome Evaluation**
- What are the differences in changes over time in regard to:
  – Correct knowledge of HIV transmission and prevention
  – Using condoms last time with different types of partners
  – Using condoms consistently with different types of partners
  – Occurrences of STIs
  – Seeking STI services
  – Seeking CT services

More detailed evaluation questions and sub-questions were included in the evaluation plan for each country evaluation.

Evaluation Approach

To fulfill the objectives and to answer the above questions, each of the MSM program evaluations used a combination of utilization-focused (Patton, 1997), mixed-method, and participatory evaluation approaches.

Data Sources and Collection Methods
All three country evaluations used the following data sources and collection methods:
- Existing program monitoring data from FHI country offices and the IAs' routine monitoring reports.
- Existing behavioral data from previously conducted large-scale Behavioral Surveillance Surveys (BSS) and/or Integrated Biological and Behavioral Surveys (IBBS).
- Individual key informant interviews with FHI country office staff, IA staff at all levels, program participants, and representatives from partner and government agencies.
- Group interviews with field staff and program participants.
- Observation sessions of program activities at offices, drop-in centers, STI clinics, and field sites such as parks, cinema halls, massage parlors, and street areas.

Reflections on the Evaluation Design

When reflecting on the three MSM program evaluations, there were two methodological features in particular that addressed the evaluation challenges discussed earlier and in our view contributed significantly to the evaluations' success. These were: a participatory approach and a mixed-method design.

Participatory Approach
Reasons to Use Participatory Methods in Evaluation
A participatory approach was used to design and implement the evaluations, mostly as a means to meet the diverse information needs of multiple audiences, and particularly of the IAs. A participatory approach was also used to gain access to and the trust of MSM organizations and their members. However, other reasons for using a participatory approach included:
- *Better quality data:* By understanding the purpose of the evaluation and having a say in how the evaluation is conducted and used, both staff and program participants may be more accessible to interviewers and more willing to provide honest information.
- *Results more likely to be used:* By having program staff and beneficiaries help plan an evaluation, the evaluation design may be more relevant and findings more credible, and, thus, the results more likely to be used (Patton 1997).
- *Greater ownership of the evaluation and program:* By learning about and having a say in an evaluation, participants may feel greater ownership of and investment in the evaluation. Similarly, as participants learn the value of their program and have a greater role in shaping its future through evaluation, participants may feel greater ownership of, and motivation to support, the program (Coghlan 1998).
- *Empowerment and acquisition of evaluative skills:* By being involved in the different stages of an evaluation, participants are more likely to undergo reflection and self-evaluation. With these skills, organizations then will be better able to "continually assess their progress toward self-determined goals and to reshape their plans and strategies" (Fetterman 1996, p. 6).

Participatory Processes and Methods Used in the MSM Program Evaluations
In the FHI MSM program evaluations, there were four major ways in which stakeholders participated:
- *Initial meetings with major stakeholders:* An evaluation consultant first met with FHI regional and country staff and then with senior IA staff. In each of these meetings, the evaluation consultant explained her independent position and participatory approach to evaluation. She then invited each organization to articulate what they wanted to learn from the evaluation and their suggestions for next steps. These initial meetings were important to ensure that a core purpose of the evaluation was to address the

information needs of the implementing agencies. Another important reason was to begin establishing trust between the evaluation teams and the individual organizations.

- *Evaluation design workshops:* A two-day evaluation design workshop was conducted with each of the implementing organizations. These workshops were attended by representatives from the head and field offices, outreach workers, peer educators, and program participants. In these workshops, participants first conducted partial Appreciative Inquiries to discover "the best of what is" in their organization and then devised simple program logic models. From there, participants decided the purpose of the evaluations, formulated general evaluation questions, and decided the methods and logistics for answering those questions.

- *Providing results throughout the evaluation:* While conducting the MSM program evaluations in Bangladesh and Nepal, preliminary evaluation findings were shared with both IA and FHI senior staff. This was done to verify and help interpret some findings, and obtain input and support for various recommendations. In addition, discussing emerging findings with senior staff meant that there were no surprises when the final results were presented.

- *Evaluation dissemination and action planning workshops:* Experience shows that merely presenting or disseminating evaluation results does not ensure their actual use (Patton 1997; Coghlan 1998). With this in mind, dissemination and action planning workshops were held with the IAs, where each workshop was slightly different depending on organizational structure and evaluation results. By country, these events included:
  Bangladesh: A one-day Evaluation Dissemination and Action Planning Workshop, in which IA staff and program participants commented on the evaluation team's major findings and recommendations, and subsequently discussed, prioritized and devised their own action plans for implementing the key recommendations.
  Nepal: A two-day Evaluation Dissemination and Action Planning Workshop, in which participants verified the findings, refined and suggested additional recommendations, and devised action plans deciding who would do what with particular recommendations.
  Indonesia: A two-day Reflections and Recommendations Workshop, in which participants verified the monitoring data, reflected on their accomplishments, strengths and weaknesses, and identified recommendations for improving program implementation.

Results of Using a Participatory Approach

Using a participatory approach to designing and implementing the evaluations had several beneficial effects. First, the evaluation teams were able to gain valuable input from the IAs and program participants in defining the evaluations' purposes, questions and data collection methods. This input made the evaluation design more relevant, realistic, and feasible. Similarly, it fostered greater ownership of the evaluation among the IAs. As one of the IA executive directors later explained, "I had ownership of the report – I saw my ideas and opinions in the report."

By using a participatory approach, the evaluation teams gained better access to program participants and better quality data. Again, as one of the executive directors said, "Earlier, [we] thought evaluation was to find something bad ... to find problems. But with participatory methods people can share ownership of evaluation ... share feelings both positive and negative. ... How else will you access beneficiaries: trust is very important. If not trust, why should they share information with you?"

By participating in the evaluation's design and dissemination, program staff and participants also felt greater ownership of the program and motivation to further support the program. For example, and as explained by one of the IA's monitoring and evaluation officers,

"[Program implementers] felt encouraged, appreciated the work . . . motivated by being involved [in] the results."

The participatory approach also resulted in the IAs using the evaluation results and recommendations. As an M&E officer explained, "Recommendations with action plans [are] more effective. [I] feel the recommendations are very good, [that they] would be good to follow." In addition, several of the IAs used the evaluation findings and recommendations to help obtain additional financial support. For example, BSWS devoted an entire issue of its quarterly newsletter in both English and Bangla languages to reporting the findings of the evaluation. The newsletter was seen not only by program staff and participants, but also by current and potential donors, and used to explain how additional efforts and funds could help address the strengths and weaknesses identified in the evaluation.

Lastly, an M&E officer described ways in which his own skills were further developed by participating in the evaluation. For example, when he now conducts site visits he has a different way of asking and incorporating questions and helping to plan for future activities. In particular, and as a result of using Appreciative Inquiry in the evaluation design workshop, he now asks about what is working well and "not just the negative."

Mixed-Method Evaluation Design and the Use of Triangulation

Another major methodological feature of the three evaluations was their mixed-method design and use of triangulation and integrated analysis.

Reasons for Using Mixed-Method Evaluation Design and Triangulation

The three MSM program evaluations used a mixed-method evaluation design, or a deliberate mix of multiple quantitative and qualitative methods. Greene et al (1989) highlight five major purposes for using mixed-method designs in evaluation:

> TRIANGULATION seeks convergence, corroboration, and correspondence of results from the different methods.
> COMPLEMENTARITY seeks elaboration, enhancement, illustration, and clarification of the results from one method with the results from the other method.
> DEVELOPMENT seeks to use the results from one method to help develop or inform the other method, where development is broadly construed to include sampling and implementation, as well as measurement decisions.
> INITIATION seeks the discovery of paradox and contradiction, new perspective of frameworks, the recasting of questions or results from one method with questions or results from the other method.
> EXPANSION seeks to extend the breadth and range of inquiry by using different methods for different inquiry components. (Greene et al 1989, p. 259)

In the MSM program evaluations, mixed-method design was used for the purposes of complementarity, development, and expansion, but primarily for triangulation. As described by Greene et al (1989):

> From its classic sources, triangulation refers to the designed use of multiple methods, with offsetting or counteracting biases, in investigations of the same phenomenon in order to strengthen the validity of inquiry results. The core premise of triangulation as a design strategy is that all methods have inherent biases and limitations, so use of only one method to assess a given phenomenon will inevitably yield biased and limited results. However, when two or more methods that have offsetting biases are used to assess a given phenomenon, the results of these methods converge or corroborate on another, then the validity of inquiry findings is enhanced (p. 256).

In guidelines for evaluating HIV/AIDS programs for most-at-risk populations (MARP), the use of triangulation is identified as an important tool, particularly as a means to assess and substantiate the linkages between program interventions and outcomes (e.g. behavior change)

and impact (e.g. rate of new HIV infections) (UNAIDS 2007). According to these guidelines, triangulation can be used to answer a number of questions, including:

- Are interventions working and are they making a difference?
- Can the observed changes in outcomes and impact indicators be attributed to the program outputs? (UNAIDS 2007, p.68)

As mentioned earlier, there often is a paucity of data regarding MSM HIV/AIDS programs. Thus, an advantage of triangulating data from multiple sources is that it can make good use of existing data and not require additional, and often expensive, data collection. Another advantage is that existing data can be used to assess and substantiate program effectiveness, by assessing process-level and outcome-level data together to address attribution and to make the case for plausible cause-effect relationships (UNAIDS 2007, p. 69).

The Use of Triangulation in the MSM Program Evaluations

All three MSM programs met the basic criteria for using triangulation (UNAIDS 2007, p. 70):

- The various programs had been in place for a sufficient duration of time to reasonably expect that changes at the population-based outcome and impact levels may be attributable to program interventions. All the programs had been operating for the recommended minimum of two to three years.
- All of the programs had been implemented with sufficient intensity and with high enough coverage to reasonably expect effects to be observed on the target MARPs at the population-level.
- Process monitoring and evaluation data were complete, of high quality and accessible for each intervention. These included program outputs, coverage estimates, and the quality of program implementation.
- The process evaluation indicated that planned program activities were, more or less, implemented as planned.
- High-quality population-based outcome and impact-level data were available that are representative of the high-risk group being targeted by interventions in the geographical catchment area of the program.

All three evaluations also used the following mix of data sources and methods:

- Program implementation or process data documenting program outputs such as number of total and new peer education contacts, number of total and new training participants, number of condoms distributed, number of educational materials distributed, and number of patients attending STI clinics.
- Existing population-based behavioral outcome data, from either Behavioral Surveys (BSS) or Integrated Biological and Behavioral Surveys (IBBS). Survey results were used to assess changes and/or differences in exposure to the interventions, correct knowledge of HIV and AIDS, using condoms both last time and consistently with different types of sexual partners, and experiencing STIs in the past 12 months.
- Qualitative data methods, such as in-depth and group interviews with program staff and beneficiaries, brief exit interviews with STI patients, and observation of drop-in center services, STI clinics, counselling services and peer educators working in cruising sites.

Despite the similarities in data sources and methods, data were analyzed and presented differently in each country evaluation. This was in large part because time series outcome survey data were available only for the Bangladesh and Indonesian evaluations, whereas Nepal survey data were for only one point in time.

Figure 1 below presents an example of triangulated data from the Bangladesh evaluation.

Figure 1: Outputs of Intervention targeting MSM and BSS MSM Outcomes, Sylhet, Bangladesh, 2001-2004
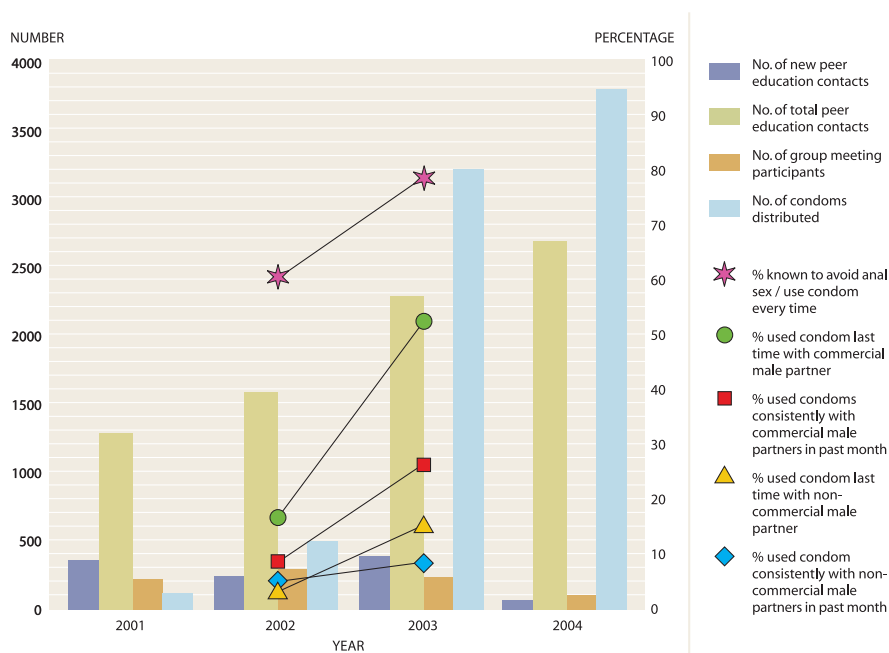

Figure 1 shows the triangulation of process data with behavioral outcomes. As the numbers of new peer education contacts and condoms distributed increased from 2002-2003, there has also been an increase in the proportion of MSM in Sylhet, Bangladesh reporting that they understand the need to either avoid sex or use condoms and to have used condoms last time and consistently with both commercial and non-commercial sex partners. It is important to note that at the time of the evaluation, program coverage was substantial and no other MSM HIV/AIDS programs were operating in the area.

Results of Mixing Methods and Triangulating Data

The presentation of the above data does not definitively prove attribution of program effects. It does, however, make good use of existing data and provide a persuasive and logical argument that the outcomes under study were the result of program activities and outputs.

Contributions to the Development and Use of Evaluation in Asia

By the nature of the HIV epidemic in Asia, and particularly in South Asia, it is entirely feasible that there will be an increase in HIV prevention and AIDS care programs in the future. The changing sources and nature of funding for HIV programs in South Asia also means that in all likelihood there will be an increase in demand for effectiveness evaluation. In designing and conducting such evaluations, donor and international intermediary agencies, local implementing agencies, and evaluators should seriously consider the use of participatory approaches and mixed-method and triangulated data designs. In some evaluation work, it is "nice" but not essential to use participatory evaluation methods. However, our experience in the FHI evaluations indicates that it was necessary to use participatory evaluation methods in

order to (a) meet the varied evaluation needs of multiple stakeholder groups, (b) gain access to and build the trust of a hard-to-reach and often wary population, as well as (c) gather quality information and (d) promote use of the results to improve program implementation. Furthermore, the FHI evaluations demonstrate not only the benefits of using participatory approaches, but also provide ways that such approaches can be implemented or adapted by other programs.

Mixed-method design and the use of triangulation also should be considered by donor agencies, local organizations and evaluators, when evaluating HIV/AIDS interventions and addressing the issue of attribution of effects. This is true particularly in light of the growing calls by some international donors and agencies for rigorous controlled experimental methodologies in the evaluation of international development programs. As an alternative to the use of expensive and ethically questionable control groups, triangulation of both process and outcome data can be used to address the thorny issue of attribution by exploring and making a case for plausible cause-effect relationships.

Lessons Learned in Evaluating MSM HIV/AIDS Programs in Asia

Despite the general benefits of using participatory approaches and mixed-method designs, there are additional issues to consider in the use of these methodological features. The first is the attitude, time and funds needed to facilitate meaningful participation throughout an evaluation. Donor agencies and evaluators need to truly want and respect the input of the local implementing agencies, not only in the evaluative information provided, but also in evaluation design, implementation and use. They also need to provide the funding and time necessary for participatory evaluation meetings and workshops.

The second lesson learned is the range of skills needed to conduct mixed-method and triangulated data designs. For example, in the FHI evaluations extensive skilled human resources were needed to: 1) pull together, verify, and tabulate the program implementation data, 2) run, analyze and interpret large survey quantitative data sets using fairly sophisticated analytical procedures, and 3) present and write-up the evaluation findings in usable formats. These diverse skills often are difficult to find or afford in the non-profit sector of resource-poor countries.

References

Bennett, S., T. Boerma, et al. (2006). Scaling Up HIV/AIDS Evaluation. Lancet 367:79-81.

Beyrer, C., S. Eiumtrakul, et al. (1995). "Same Sex Behavior, Sexually Transmitted Diseases and HIV Risks among Young Northern Thai Men." AIDS 9(2): 171-6.

Centers for Disease Control and Prevention (2006). HIV Prevalence among Populations of Men who have Sex with Men – Thailand, 2003 and 2005. MMWR 2006; 55: 844-848

CDC-GAP/USAID-RDMA-FHI/APD (2005). Strategizing Interventions among MSM in the Greater Mekong Sub-region (GMR). Bangkok, Centers for Disease Control and Prevention/Global AIDS Program – United States Agency for International Development/Regional Development Mission – Family Health International/Asia Pacific Department.

Choi, K.H., H. Liu, et al. (2003). "Emerging HIV-1 Epidemic in China in Men who have Sex with Men." Lancet 361 (9375): 2125-6.

Coghlan, A. (2006). Evaluation of the FHI/Indonesia Aksi Stop AIDS (ASA)-Supported MSM Implementing Agencies. Unpublished report.

Coghlan, A. (2005). Evaluation Plan for Family Health International/Asia and Pacific Division's Men Who Have Sex with Men Programs. Unpublished Report.

Coghlan, A. (1998). *Empowerment-oriented Evaluation: Incorporating participatory evaluation methods to empower Ugandan communities to prevent HIV/AIDS*. Ph.D. Dissertation, Cornell University, Ithaca, NY.

Coghlan, A. and Khan, S.I. (2005). Bandhu Social Welfare Society – Family Health International MSM Program Evaluation. Unpublished report.

Colby, D.J. (2003). "HIV Knowledge and Risk Factors among Men who have Sex with Men in Ho Chi Minh City, Vietnam." *J Acquir Immune Defic Syndr* 32(1):80-5.

Fetterman, David M., Shakeh J. Kaftarian, and Abraham Wandersman (eds.) (1996). *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability.* Thousand Oaks, CA: Sage Publications.

Girault, P., T. Saidel, et al. (2004). "HIV, STIs, and Sexual Behaviors among Men who have Sex with Men in Phnom Penh, Cambodia. *AIDS Educ Prev* 16(1): 31-44.

Go, V.F., A.K. Srikrishnan, et al. (2004). "High HIV Prevalence and Risk Behaviors in Men who have Sex with Men in Chennai, India." *J Acquir Immune Defic Syndr* 35(3): 314-9.

Greene, J.C., Caracelli, V.J., and Grahm, W.F. (1989). Toward a Conceptual Framework for Mixed-method Evaluation Design. *Education Evaluation and Policy Analysis*, 11(3), 255-274.

Johnson, W.D., L.V. Hedges, et al. (2002). "HIV Prevention Research fro Men who have Sex with Men: a systematic review and meta-analysis." *J Acquir Immune Defic Syndr* 30 Suppl: S118-29.

MAP (2005). Male-Male Sex and HIV/AIDS in Asia. Available at: http://www.mapnetwork.org/docs/MAP_&M%20Book_04July05_en.pdf.

Naz Foundation International (2003). Model for MSM Sexual Health Interventions. Naz Foundation International.

New ERA and Coghlan, A. (2006). Evaluation of the FHI-Supported Behavior Change Intervention of the Blue Diamond Society. Unpublished report.

Parker, R. K.S., Aggleton P. (1998). "Conspicuous by their Absence? Men who have Sex with Men (MSM) in Developing Countries: Implications for HIV prevention." *Critical Public Health* 8(4):329-345.

Patton, M.Q. (1997). *Utilization-Focused Evaluation: The new century text, 3rd edition.* Thousand Oaks, CA: Sage Publications, Inc.

Pisani, E., P. Girault, et al. (2004). "HIV Syphilis Infection, and Sexual Practices among Transgenders, Male Sex Workers, and Other Men who have Sex with Men in Jakarta, Indonesia." *Sex Transm Infect* 80(6):536-40.

Preston, D.B., A.R. D'Augelli, et al. (2004). "The Influence of Stigma on the Sexual Risk Behavior of Rural Men who have Sex with Men." *AIDS Educ Prev* 16(4):291-303.

Sittitrai, W. P. Phanuphak, et al. (1994). "A Survey of Thai Sexual Behaviour and Risk of HIV Infection." *Int J STD AIDS* 5(5): 377-8.

UNAIDS (2007). *A Framework for Monitoring and Evaluating HIV Prevention Programs for Most-At-Risk Population.* Geneva: UNAIDS.

van Griensven, F., P.H. Kilmarx, et al. (2004). "The Prevalence of Bisexual and Homosexual Orientatin and Related Health Risks among Adolescents in Northern Thailand." Arch Sex Behav 33(2): 137-47.

van Griensven, F., S. Thanprasertsuk, et al. (2005). "Evidence of a Previously Undocumented Epidemic of HV Infection among Men who have Sex with Men in Bangkok, Thailand. *AIDS* 19(5): 521-6.

Wilson, P.A. and H. Yoshikawa (2004). "Experiences of and Responses to Social Discrimination among Asian and pacific Islander Gay Men: Their relationship to HIV risk." *AIDS Educ Prev* 16(1): 68-83.

Wulfert, E., C.K. Wan, et al. (1996). "Gay Men's Safer Sex Behavior: An integration of three models." *J Behev Med* 19(4): 345-66.

Evaluators, donors, program managers and international agencies sometimes seem to care only about methodologies, methods and evaluation designs. Techniques are invisible, relegated to the status of poor cousins. Yet choosing the right technique or tool for the right reason in the right context is one of the most important skills an evaluator can possess. Poor or inappropriate technique can ruin the best method. A bad questionnaire? Useless, misleading data no matter how fancy the evaluation design.

So the editors were delighted to have an article that purely focused on technique. The tools described below were developed in Asia specifically as a way of making the invisible (attitudes, values, perceptions) visible, for generating debate about things that are often difficult to discuss. But the ideas that underpin them mean that, with care, they can be adopted and adapted in other contexts for other purposes.

New Participatory tools for measuring Attitude, Behavior, Perception and Change

Dr Ravi Jayakaran

Project and Program evaluations are finally gaining greater recognition for the value they add to the quality of programs. This is also taking root in South Asia, and professionals are beginning to look at carrying out evaluations of their programs, not only because it is a part of 'donor requirement' but because they genuinely feel that it will benefit the program to improve quality and move towards becoming more sustainable. Evaluation feeds into knowledge management that in turn enables the project leadership team improve the quality of the program. This can be done through tracking aspects of the program that have not been effective or that require modification. Evaluation thus becomes a powerful tool to improve the effectiveness of a program, and also enable the efficient use of resources in the program. However, evaluations have often been perceived as being cumbersome, protracted and time and cost consuming. Also significant dimensions of the program such as attitude and behaviour change are difficult to measure, and this has discouraged some project holders from attempting such an evaluation. The focus of this article is to describe some participatory tools that can be used for measuring attitude and behaviour change. Participatory tools are easy to use, and can be learned by project staff without having to draw on the services of costly evaluation experts.

It is necessary to begin with a hypothesis when designing a program. Someone has said that "vision is 20/20 hind sight". This essentially means that after the incident has occurred we can see clearly why it went wrong. However, at this stage it is too late to change! Timely evaluations therefore can help us make corrections while the project is still underway, and to even understand the relevance of the program intervention and determine if it is likely to be sustainable; so that necessary and appropriate changes can be made during the remaining part of the program.

Attitude, behaviour, perception and change in the community are some of the most significant 'intangibles' that social scientists and development professional have sought to measure while assessing the impact of development interventions. It is easy to measure an action when it has been completed, but it is difficult to measure its process. Similarly it is hard to measure to what extent a community is open to a new idea or resisting it. However knowing this is important to make mid-course corrections to the program or intervention Similarly, when a program has been operational for some time, how does one assess its impact? Or find out how well it has progressed when there are no tangible outputs that can easily be measured.

How does one, essentially, measure attitude, behaviour and perception? Frustrated with this problem and to overcome these difficulties, the author developed several new participatory tools and field tested them. These participatory tools were developed in ways that they would be simple, easy to learn and use. Besides being simple the tools also enable considerable discussion and dialogue, thus facilitating the natural outcome of the process, namely to explore ways to modify, improve and correct existing programs.
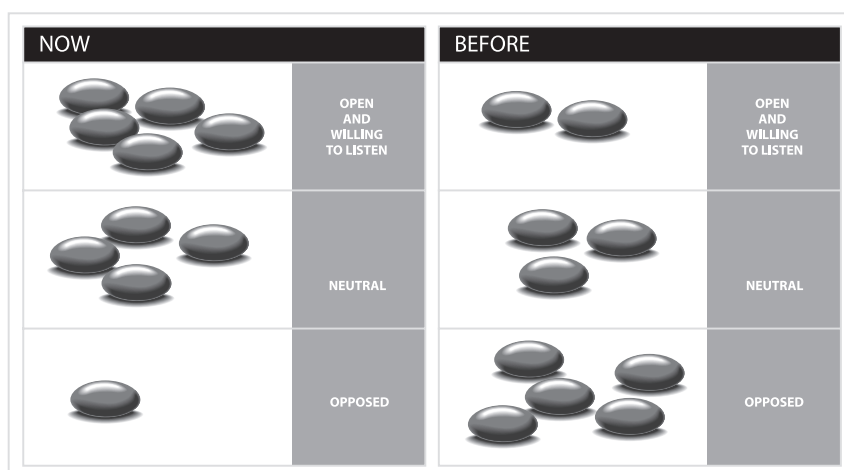
The rest of this article profiles some of these new techniques along with examples of situations in which they have been used. Also given under each exercise are the lessons learned, and new insights drawn from the using it as a participatory evaluation tool.

Measuring Attitude:

The following are some of the exercises developed for measuring community attitude to an intervention or program:

1. O.T.C. (Openness to Change) Index
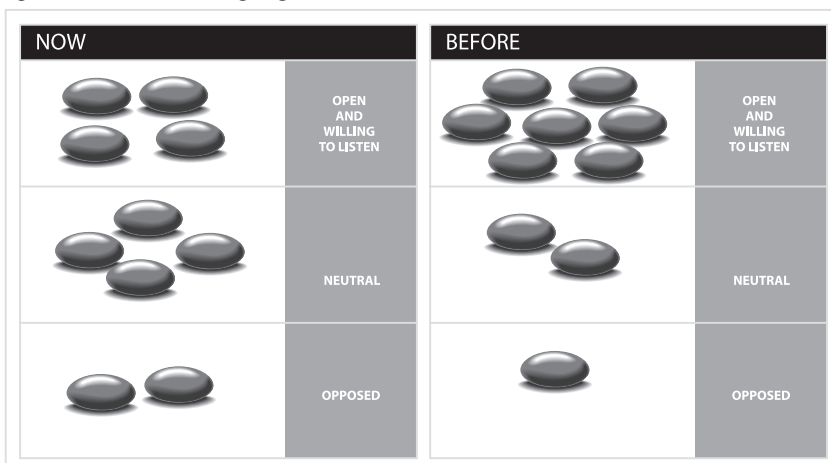
Figure 1: Openness to change

The OTC exercise is used to find out the attitude of the community or group to new ideas or to a program that requires change. The exercise is done using the TST (Ten Seed Technique). A focus group consisting of 8 to 10 people from the community is asked to think of the 10 seeds as representing the total community in question. They are then asked to distribute the seeds into 3 boxes as seen under the section under "Now" in *Figure 1*. The distribution of the seeds into three groups of those who are "open & willing to listen", those who are 'neutral' and those who are 'opposed' and the rationale behind it is then discussed with the group. After the discussion, the group is asked to distribute the seeds to show the status of how things were at the early stages of the project. This provides a basis for discussion, as the community can be asked about the reasons why there was this change of attitude in the community. For instance, the following reasons were offered by group members for the shifts in attitudes of their particular community to the changes being introduced:

- Because the change agent took time to interactively explain to the community the entire program
- The change agent took time to build relationships and interact with the community
- and the change agent was flexible and willing to make changes in the program to accommodate community needs

Lessons learned: The reasons the community attributed positive change were essentially related to the specific change agent from the NGO who interacted with them, built relationships with them, explained things to them, and was willing to be flexible to accommodate the changing needs of the community. Getting to know this information helped the NGO understand what was appropriate in that particular context from the community's point of view, and they were then able to ensure that relationship building was an essential part of future expansion. This learning affected the entire program as other change agents were encouraged to observe and learn from the performance of this particular change agent.

2. A.T.C.A. (Attitude to Change Agents) Index

Figure 2: Attitude to change agents

It is clear from the example given above that the role of the change agent can be very crucial in the development process; which is heavily dependant on the relationships that exist with a community. In the ATCA exercise shown in *Figure 2*, the NGO involved was attempting to find out what went wrong in a particular target area where they were working. After using the TST, the seeds were distributed into the three boxes by a group of village leaders and decision makers (showing the current situation under 'Now'). After discussions, the group was asked to show the profile of how things were at the time when the project first started. The difference in the community attitude was very radical. It became clear that the change agent responsible for this area had aggravated the community ,turned them off, and in some cases even made them 'opposed' to the NGO. To find out where failure had taken place the focus group was asked to share why this change in attitude had occurred. The reasons were almost entirely related to the way the change agent lived while he was in the community, often confronting them and being partial in the way he behaved with some people and neglecting some of those whom he was supposed to serve. The exercise helped the NGO concerned to take appropriate action by removing the change agent from the area (and ultimately from the project).

3. E.T.C. (Enablers to Change):

This exercise is usually done when projects and programs that have been successfully completed or achieved substantial progress by way of arriving at significant milestones in their life cycle. The focus group for this exercise consists of team leaders, representative leaders (from the project or program) and other stakeholders. The group goes through a quick review of the progress of the project during its life time up to that point of time, and then carries out the exercise shown below:

Figure 3: Enablers & Barriers To Transformation

Enablers	Barriers
0000000	000

Using the TST, the group is asked to distribute the seeds in the two boxes to show the ratio between the "enablers" and "barriers" to transformation that they consider that the project has experienced. The diagram ensuing from this placement is then discussed (as with all the other preceding exercises), till all members of the group are actively involved and contributing to the discussion. After this is done, each of these is further discussed to find out what the enablers are (*Figure 4*) and what the barriers are (*Figure 5*).

Figure 4: Enablers for Facilitating Transformation

Enabler	Intensity
Good model of the staff and the Core Group	000
Helping each other (between the community and the Core Group) and also good communication between the Core Group and the community	00
Poverty reduction	000
Openness to Change	00

4. B.T.C. (Barriers to Change)

The "barriers to change" exercise is also conducted to explore further the "barriers" and find out what these are. The following is the way the seeds were distributed:

Figure 5: Barriers for Facilitating Transformation

Barrier	Intensity
Less people involved in village meetings and less participation	00
Project still unable to prevent some diseases in humans and animals	00
'Partyism' (tendency to form small cliques)	0
Lack of capital for business	000
Dependence on the Project & outsiders	00

Lessons learned: Getting this information about the enablers and barriers to transformation helps understand what is relevant to the context and provides insights into how the programs can be further improved to achieve greater impact.

Measuring Behavior:

1. R.T.P. (Response to Programs) or P.I.A. (Program Impact Assessment)

The RTP or PIA exercise is used for assessing the response to a particular program to see how the community has been impacted. To do this exercise, one uses the PDM (Project Design Matrix) of the project and creates the impact matrix that is seen in *Figure 6*. On the left side of this matrix is the list of all the impact areas that the project is intended to effect. In front of each impact area there are two columns, one to show what the "input has been till now" and the other showing "what the matching output was" for that particular impact area. The Focus group is asked to use the TST and distribute the seeds for each impact area between these two columns.

The focus group for this exercise constituted leading members of the community from the target area and the leadership team of the project along with other local stakeholders. As they consider how the seeds should be distributed, the group is advised to concentrate on impact on the community rather than on actual program activities.

The final details of the exercise enable an analysis of the change in behavior of the community as a result of the various programs in the various impact areas. The RTP/PIA analysis of the project result from an actual analysis of a project is seen in *Figure 6*.

Figure 6: Program Impact evaluation PH -2 Project (Input–output ratio for different program components)
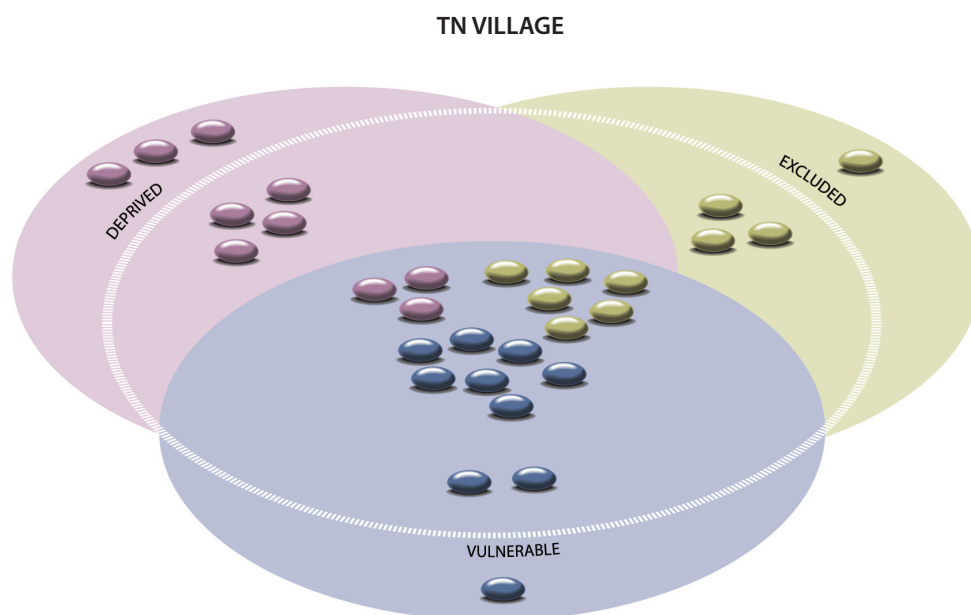
Activity		EFFICIENCY	
		Input	**Output**
Efficiency of field activities based on INPUT–OUTPUT ratios	EDUCATION	00000	00000
	HEALTH & NUTRITION	000000	0000
	VALUE CHANGE	000	0000000
	WATER RESCOURCES	000	0000000
	ECONOMIC DEVELOPMENT	0000000	000
	ORGANIZATIONAL DEVELOPMENT	000000	0000
FINANCE		000000	0000
MANAGEMENT		0000	000000
PLANNING		0000	000000
SATISFYING EXPERIENCE		0000	000000
This exercise was conducted with the team to enable them to judge for themselves how their efforts are paying off, and also how they can maximize their efforts to get better results in the areas where they have already acheived maximum efficiency. – Ravi			

Lessons learned: In this example the best results were found with impact in the area of value change and creating water resources for the community. On the other hand, the programs for economic development and OD required more efforts than the results they produced. Carrying out this exercises enabled the group understand which programs gave the best results. They were also able to determine which program impact areas were not as effective. Ultimately, as with all such "visual" output creating exercises, the "visual" (output above in *Figure 6*) is interviewed. This is done by looking at various aspects of the output and asking the group in general, "why is this ratio high or low?" Individuals in the group, who are familiar with why this is so, respond giving reasons. The answers provide clues on ways to make changes and these are noted by the management for taking appropriate action.

2. D-E-V (Deprivation-Exclusion-Vulnerability) Index:

This exercise is used to assess to what extent children are "at risk" or "safe" within a community. The focus group for this exercise consists of community leaders and other stakeholders and they assess the status of the community to see to what extent children are **Excluded, Deprived** and **Vulnerable** in that particular community. This can be done at a village level, or a community? level, or a district level. Depending on the level at which it is conducted, group members can be either village leaders, or commune leaders or district leaders. It is however mandatory that the focus group knows the area they are assessing very well. The most accurate results are obtained when the exercise is done at a village level. However it can also be done at macro level (say at a province or state level) to initially identify the districts that are vulnerable. Then within these districts, identify the communes that are at risk, and finally zero in on the villages that are at risk. Ultimately, the families at risk within the village can be identified. (*Figure* 7 below)

Figure 7: Some profiles from Cambodia
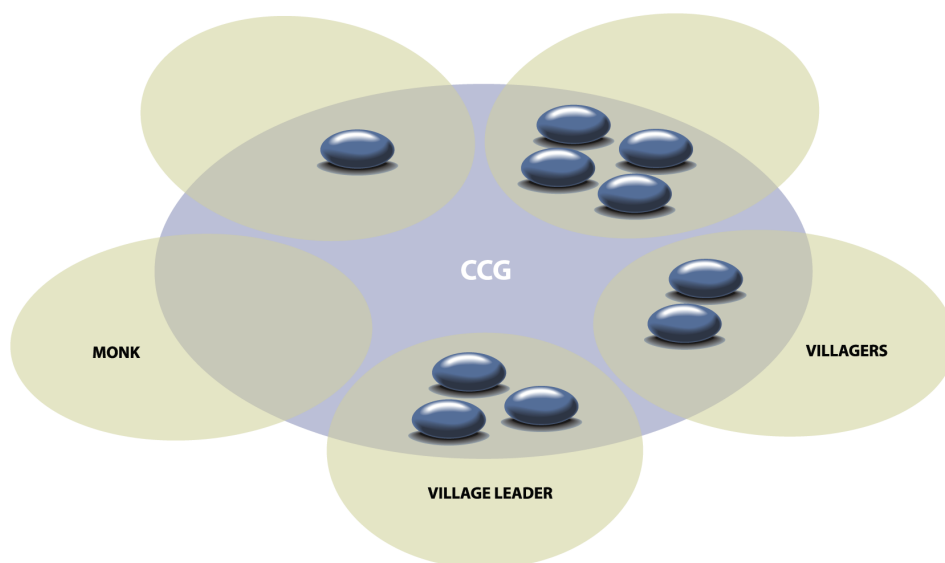

TN VILLAGE

Each of the three sets of 10 seeds (three different colors) are representative of each of these dimensions (Deprivation, Exclusion and Vulnerable) and the group is asked to distribute the seeds into three groups. Starting with identifying the excluded, the outer most part is represented by the % of the population that is excluded, in the middle for those partially excluded, and within the three overlapping circles for the proportion which are safe. This is done similarly for the "Deprived" and "Vulnerable" aspects too. In the case above for "TN" village (*Figure* 7) the DEV Index for Risk is "5" and for safe is "16". When carrying out this exercise for assessing a group of villages to see which ones are safer or more at risk, an external member who remains constant for each village is added and this person facilitates the process, so that there is an element of uniformity and comparison. Villages can thus be ranked for priority in development interventions. The tool can also be used for development intervention planning with the focus group once the 'visual' is ready. The exercises can be facilitated at

different stages of the life cycle of the project to see, for instance, if there is any reduction in the number of children at risk. Once "at risk" villages are identified, special attention can be given to them for improvement and also for prevention of trafficking.

3. D.M.P. (Decision Making Profile):

Figure 8: Village Code: KS-2 (Current decision making profile)

The Decision Making Profile exercise is used for finding out how participatory the decision making process in a village is and especially to see if any key people from the village have been left out of the decision making circle. The tool also serves as a quick reminder and corrective measure for mid term or mid course corrections in strategy for better decision making that allows for better quality and greater ownership.

Lessons learned: In the example above, the evaluation identified and brought to light why the monk at the local pagoda was not cooperating with the team spreading HIV/AIDS awareness in the village. The monk had been left out of the decision making process when it came to planning important activities in the village, and as a result, when the team went to ask him for permission to organize meetings at the pagoda, the monk didn't cooperate. After understanding the situation through carrying out the DMP exercise, the group was able to take the result to him (*Figure 8*) and apologize for failing. The result was immediate, with the monk immediately becoming friendly and cooperative.

4. D.P.P. (Default Priority Profile) [2]

Figure 9: Participatory Evaluation of IH -2 project

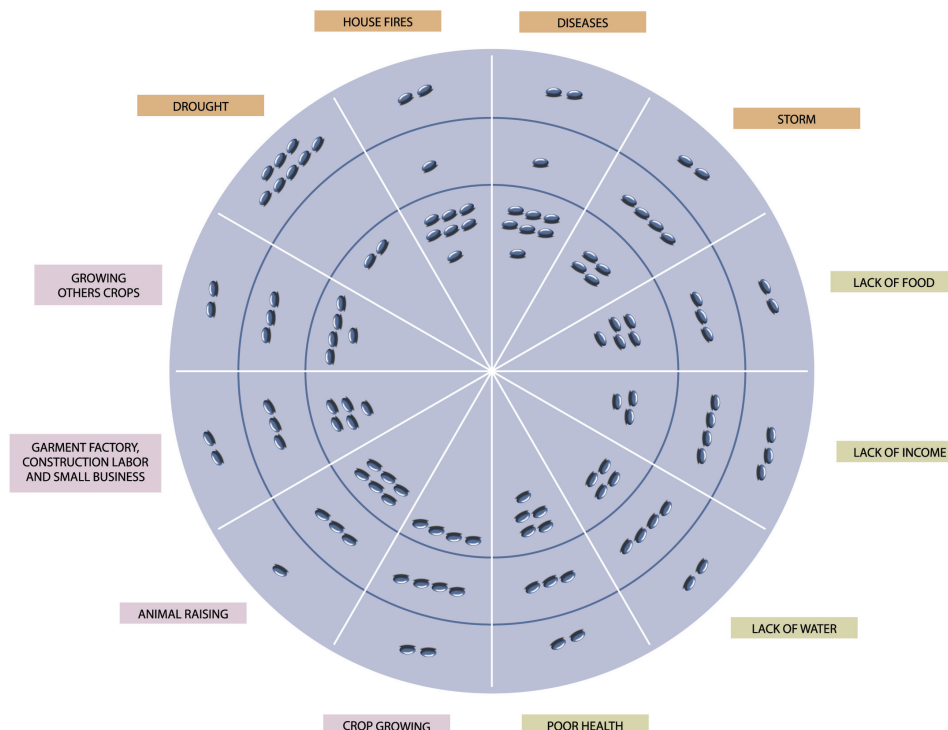| Areas of Priority	D.P.P. Index	Type of Orientation	Focus	Project Coordinator's Remarks
Strategy		Macro impact	Overall Ministry orientation	Aligning with the different FBQs Fund Sourcing Practical Capability of project partner Srengthening of IGPs
Management & Administration		Efficiency	Audit conscious orientation	Participatory management of the project Strenghtening partnership of the project (VW and Project Partner through project management board)
Program Implementation & Supervision		Effectiveness	Micro managing orientation	Delegation of the monitoring activities to the project staff and volunteer leaders
Grass roots/Ground level involvement		Relationship	Grass roots level orientation	Regular dialogue and consultation with beneficiaries

The DPP exercise can be used at various levels to assess an individual leader, a project, a program, a country office a regional program or an international operation. There are four levels of involvement as seen in *Figure 9* above for maximizing impact. These are: at the strategy level, at the management and administration level, the program implementation and supervision level, and finally at the grass roots level. The type of emphasis given to each of these will result in a particular result as seen in the figure above. Experience has shown that the ideal overall ratio for maximum response is 4:2:2:2. In the example above from a program evaluation, there was an overemphasis on the program implementation and supervision level, resulting in a compromise on time spent and emphasis given to grass roots level relationship building. The tool can also be used for a program leader to carry out a "self-examination" with "360 feedback" (feedback from supervisor, peers and direct reports) to assess where she of he is at with their DPP. After this, they can make corrective changes in their emphasis for greater effectiveness, supported by an action plan to overcome shortcomings. The DPP can also be used as an evaluation tool to see the emphasis that a project or program has placed on the 4 areas. This can be done with a focus group of key stakeholders, including the project or program leader. The example above is from a program evaluation.

Measuring Perceptions:

 1. **W.W.V.A. (Wholistic Worldview Analysis)**

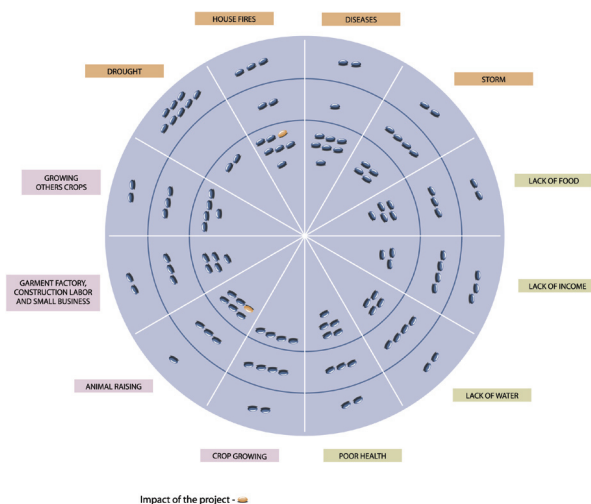*Figure 10:* WWVA of RP Village

The WWVA exercise is now being used in several countries in the Greater Mekong sub-region to understand community survival strategies and for planning development interventions. The WWVA exercise draws on the information emerging from three other exercises not described in this article, namely a livelihood analysis, a problem analysis and an uncertainty analysis. The output of these exercises form the main component of the WWVA exercise and the focus group (constituting of village leaders) is asked then to use the ten seeds placing them for each component in one of the three parts of the concentric circles. The innermost circle represents the aspects of the survival strategy that the community controls. The second outer circle represents the aspects of the survival strategy (for that component – health, water, food etc) for which the community is dependent on outsiders. The outermost circle represents the areas of the survival strategy that are completely out of the control of the community as well as outsiders. When completed, the diagram is read as follows:

  • The outermost circle represents the Community's immediate vulnerabilities
  • The component with the highest number of seeds in the outermost circle, thus becomes the first priority for development
  • The innermost circle represents the community's strengths
  • The middle circle of course represents the second level of vulnerabilities that the community has (depending on what extent outsiders on whom they are dependent are exploiting them).

The tool has served as an effective tool for evaluating, discussing and redesigning programs. Redesigning is done by identifying each area of vulnerability, carrying out a stakeholder analysis for those who can bring about change, analyzing the extent to which they can make a difference, and then planning for action based on what is required for each to build capacities and reduce vulnerabilities (for details see detailed article on "WWVA as a development intervention tool" [3]). The second diagram here shows an example of how the WWVA diagram can be used to identify areas on which the project is concentrating. This is done by distributing 10 seeds again to show where the project has concentrated. An additional paper is kept outside of the WWVA diagram for the group to place seeds that are not related to the WWVA. In the above case there were only 3 seeds within the diagram and 7 seeds outside! (it is easy to see how some of a project or programs activities end up being outside of the community's actual needs!!)

Impact of the project -

Lessons learned: Ideally, this exercise should be done in the planning phase of the program. However, if this didn't happen it is advisable to use this tool as an evaluation tool as soon as possible so that mid course corrections can be made to avoid program failure. This tool is particularly useful in identifying if the program is the result of a "supply push" (donor push) or a "need pull" (community need oriented program).

2. U.O.T.P. (Usefulness of Training Program)

This exercise is done in two stages with the staff of a project or program who have been through various training programs: The first part being an analysis of the usefulness of the training program as a part of the overall training programs.

Figure 11

CHART training	Other training in general skills
000000	0000

After this, give the project staff another ten seeds and ask them to group these to show how that particular training was (very useful/somewhat useful/not useful). In this case they grouped it as seen in *Figure 12*.

Figure 12

Usefulness of the CHART training	Impact	Remarks
Very useful	0000000	• The ability to facilitate brainstorming • Facilitation skills in general • Communication skills
Somewhat useful	000	• Some things were not applicable to Khmer culture
Not useful		

After the seeds are distributed , ask the group to give details of why they grouped the seeds in the way they did, and also ask them to elaborate what aspects of the training was relevant or not relevant.

Lessons learned: Besides lessons learned on the usefulness or the lack thereof , of some training programs, this participatory evaluation exercise can also then be used with former trainees to modify and improve future training programs.

Measuring Change
1. P.A.P.A.(Progress Against Plan of Action)

Figure 13: Progress Against Plan of Action PH-2 Project

Impact areas		Achieved		Yet to be achieved	
		Seeds	Remarks	Seeds	Remarks
I	Food sufficiency/ Economic development	●●●●●●	Loan opportunities Infrastructure support	●●●●	Strengthening MED Food always in the house Food during the lean months
II	Health	●●●●●●●●	Strong linkage with RHU & private clinic Training local health workers	●●	Advocate community based health service
III	Basic housing	●●●●●	Infrastructure support Housing assistance (construction and repair)	●●●●●	Land and housing
IV	Education	●●●●●●●●	Provision of educational assistance (Tuition, allowance, supplies and uniform)	●●	Provision of other educational facilities such as books, desks, chairs
V	Community value change	●●●●●●●●	Spiritual nurture activities Bible cell groups CE materials	●●	Partnerships with different action groups in the area
VI	Organizational development	●●●●●●●	Capacity building of leaders Technical assistance (planning and policy making)	●●●	Organizational structure review (restructuring)
VII	ADP management	●●●●●●	Mentoring Support to project partner	●●●●	Staff development competencies (technical, design formulation and strategy development)

The PAPA is useful both as a participatory program monitoring and participatory program evaluation tool to assess progress. During an evaluation, this is used first and the project or program team is asked to use the TST to assess progress. Both what has been achieved and what still remains to be achieved are listed as shown in *Figure 13*. The team is then asked to give details under each impact area, of specific achievements, and also on the other side, areas which did not achieve what was expected. This output of this exercise can then be used as the framework of the evaluation design.

Lessons learned: This tool has been extensively used by the author for program evaluations in the Philippines, in Myanmar and in Cambodia. Currently, the tool is also being used for a major evaluation of an HIV/AIDS program in Cambodia.

2. V.C.I. (Value Change Index)

Figure 14: (VCI) Value Change Index

Every program seeks to incorporate some values as part of the intervention whether these be in terms of best practices for health and hygiene or behaviour standards related to dealing with others in the community, by way of equality for gender, care of the vulnerable (disabled, widows, children) or relations (physical and otherwise). The VCI can thus be used for assessing to what extent those were achieved, and the community can also be asked to use its own criteria for assessing what has changed. In the example above, the seeds that fall into the blue overlapping circle represent changes that have already become visible or evident in the community, with the number of seeds representing the approximate % of the population that exhibit this change. The total of seeds in the blue overlap circle represents the VCI for that village – in this case "48".

Lessons learned: Again, by having an external facilitator who remains common for a group of villages, it is possible to list the villages according to descending order of performance. The villages that perform high and those that perform poorly can be studied for "lessons learned" and will form the basis for "contextual improvement of the program" – a sort of "in-house-on-the-field" self assessment and learning program. This also opens opportunities for the emergence of new perspectives for future development.

3. **E.E.I.R.S.(Efficiency, Effectiveness, Impact, Relevance and Sustainability) for overall Quality & Performance of a project:**

Figure 15: Project overall performance in Quality (EEIRS)

E.E.I.R.S. Criteria	Overall performance	
	Achieved	**Yet to be achieved**
Efficiency	••••••	••••
Effectiveness	••••	••••••
Impact	••••••	••••
Relevance	••••••••	••
Sustainability	••••••	••••

The EEIRS tool was developed by the author based on a program supported by a Japanese international NGO that wanted to assess the project management and performance based on efficiency, effectiveness, impact, relevance and sustainability. The format seen in *Figure 15* was first used to get an overall picture from the project staff and key stakeholders who formed the core focus group. This was then followed up with them to go into more details, by elaborating each of the five criteria, and asking for details of what was achieved and what remained to be achieved (similar to PAPA in *Figure 13*). This then became the broad framework on the basis of which the program was evaluated. The outcome was that it was relevant to the donor agency as well as the project staff who got to see how they were performing in all the five areas and were able to make mid course corrections in their plans accordingly.

Lessons learned: After developing this tool for the Japanese NGO, the author has subsequently used this participatory tool for several NGO program evaluations. The EEIRS tool is also currently being used to evaluate the overall performance of a major HIV/AIDS awareness program in Cambodia.

Additional Value of the Tools

The additional advantage with these participatory tools is that they are flexible and can be adapted to different situations and different needs. In recent months similar tools have also been used for a variety of pilot assessments and participatory assessments for the Asian Development Bank to explore the impact of Regional Economic Integration on Poverty in the Greater Mekong sub-region and for assessing the environmental impact of the economic corridors on a particular commune in the SW part of Cambodia called the Cardomom mountains. Some very interesting aspects of attitude, behaviour and perspectives of both the poor and the rich (who impact them) have emerged from these studies.

Recommended further reading

1. Participatory Poverty Alleviation & Development: Ravi Jayakaran, 2003/2005, Mekong Institute Foundation, Khon Kaen University, Khon Kaen, Thailand 40002
2. Facilitating Small Groups: Ravi Jayakaran, 2003/2005, Mekong Institute Foundation, Khon Kaen University, Khon Kaen, Thailand 40002
3. TST: Ten Seed Technique: Dr. Ravi Jayakaran (see details below for download instructions).

References

1 TST: Ten Seed Technique: By Dr Ravi Jayakaran
   http://www.fao.org/Participation/ft_show.jsp?ID=1981, (to download the TST e-book)
2 D.P.P: Default Priority Profile. Covered in Chapter 16 of *Facilitating Small Groups*: CD Rom available through Mekong Insititue Khon Kaen University, Thailand 40002
3 The WWVA tool as an evaluation and Design tool:
   http://www.iied.org/NR/agbioliv/pla_notes/current.html

PART THREE – FILLING IN THE PICTURE

Some years ago John Kingdon observed that things happen quickly in civil society when three normally separate streams converge; the problem stream, the political stream and the policy stream. The problem stream is the list of items of current concern to the public, popular media and lobbyists. The political stream is the matrix of potential political pay-offs elected officials might reap from enacting various strategies. The policy stream comprises the "primeval soup" of policy solutions developed by experts, such as researchers and government bureaucrats. Two streams converging isn't enough - it has to be all three. The convergence cannot be brought together deliberately, life is way too complex for that degree of control. But skilled social entrepreneurs know intuitively when this convergence happens and seize the moment. Have Kingdon's "Garbage Can Model" in mind as you read this intriguing article.

Reference: Kingdon, J. W. (1995). Agendas, alternatives, and public policies *(2nd ed.). New York: Longman.*

Civil Society Partnership in Promoting an Evaluation Culture in the Development Process – experience of the Sri Lanka Evaluation Association (SLEvA)

P. Indra Tudawe, Ph.D
Mallika R. Samranayake

The Sri Lanka Evaluation Association (SLEvA) is a voluntary civil society organization established by an interested group of professionals to promote evaluation as an integral element in the development process in Sri Lanka to contribute to better management, transparency and accountability both in the Governmental and Non-Governmental sectors.

The need for a professional organization focusing on evaluation

The primary need for an organization such as SLEvA, stemmed out of the mindsets and practices that did not enable the optimum use of resources in the country. Existing evaluation practice, as an assessment and a learning process left much to be desired. This situation was caused by many factors and significant among them were as follows:

- Evaluations were "donor" driven for funded projects and therefore had a low acceptance and low utilization of findings by the implementers and decision-makers. Nor did evaluation create an impact on or generate a sense of ownership in the minds of the ultimate beneficiaries. Such evaluations were considered as "outsider's views" and were not made use of to improve performance. Stakeholders were not partners in the process of evaluation and treated evaluation only as a learning exercise.
- Evaluations were considered to be "fault finding missions" rather than inputs for program improvement. Various stakeholders in the development process were gravely concerned about the failure to understand the discipline of evaluation as a mechanism to identify information gaps, promote lessons learned from experiences to feed in to new projects or programs in order to avoid bottlenecks, delays in implementation, and unplanned negative outcomes or impacts.
- No institutional mechanism in the form of a professional body to further the discipline of evaluation. Consequently there was no-one to promote evaluation, encourage capacity build-ing towards developing a critical mass with the potential to influence the establishment of evaluation standards. This situation resulted in inadequate standards for quality evaluations and the absence of a sense of collective commitment towards achieving development goals.

How SLEvA came to be – a historical sketch

The evolution of SLEvA as an Association and a professional body took place against the backdrop of the above situation. Like many other professional bodies, SLEvA came into being when a group of like minded individuals formed an interest group to foster, nurture and develop a concept towards a common goal with the view to contribute to the betterment of the society. In this case a group of professionals committed to the promotion of evaluation as a development discipline, who met informally, discussed and developed the concept and the mechanism to operationalise evaluation.

The group formulated a constitution specifying the aims and objectives. Within an year of the initial meeting of the group SLEvA was formally established in September 1999, with an inaugural meeting attended by the initial group and invitee professionals of different disciplines, engaged in research, assessments and evaluations. At the official inauguration of SLEvA, a total of 35 persons registered as members of SLEvA, and membership drive since then has increased the membership to beyond a hundred.

The catalyst of SLEvA, among the initial group of individuals, was the then incumbent Monitoring and Evaluation officer of UNICEF – Sri Lanka. The group came from various disciplines such as monitoring and evaluation, socio-economic research, medicine, nutrition, geography, sociology, economic development, accountancy and chemistry. In terms of organizational representation there were individuals from UNICF, UNDP, Government research organizations and Ministries, Graduate and Postgraduate Faculties of Universities, Private Sector and Non-governmental Organizations.

The Sri Lanka Evaluation Association is a non-profit, non-governmental organization registered with the Ministry of Social Services as per the regulations of the country, as a voluntary professional organization. Its role is to encourage greater accountability and transparency, across all sectors, towards a culture of evaluation for improved wellbeing of the community.

Process followed by SLEvA to promote an evaluation culture

SLEvA has a clear definition of "evaluation culture". Evaluation culture within Sri Lanka is defined as internalizing and incorporating evaluation in program design and implementation as an input to improving management of projects and programs. This input enables corrective action where resources were under or inefficiently utilized and to maximize benefits to and impact on the target population.

SLEvA was fortunate to have a number of international partners who assisted in furthering the aims of the Association. The UNICEF primarily along with the UNDP and the GTZ were among the key agencies that assisted with capacity building and studies for promoting evaluation standards.

SLEvA formed early links with the Ministry of Plan Implementation (MPI) – a key Ministry in the Sri Lankan government. This Ministry is mandated to coordinate, monitor and evaluate projects/programs and is the focal point for the Millennium Development Goals in the Government promoting Results Based Planning and Management. SLEvA considers partnership with this agency critical to promoting an evaluation culture. Being the stakeholder handling funds owned and derived from the population at large, the Government's accountability to the public surpasses any other stakeholder in that respect. Therefore, SLEvA thought it appropriate from the beginning that the Government needs to be brought on board in promoting an evaluation culture in the development context. Such a partnership could also indirectly influence the non-government and private sector interventions, particularly in capacity building and creating opportunities for information sharing.

Four key focal areas were identified after a series of discussions among the membership and other stakeholders:

- Assisting policy formulation
- Capacity building
- Information dissemination
- Networking

Together they would help promote an evaluation culture and further the Results Based Management focus of the Millennium Development Goals.

Towards Policy Formulation

A significant achievement of SLEvA has been the Draft National Evaluation Policy for the Government of Sri Lanka. The Government of Sri Lanka believes that evaluation, a powerful tool in results-based management, is not adequately utilized in development programs. Thus the need for a national evaluation policy was highlighted at the SLEvA Conference in January 2003 by the chief guest, Secretary Ministry of Finance and Policy Development and Implementation. He requested SLEvA, as an independent professional body, to prepare a draft National Evaluation Policy (NEP) document.

The Association formulated the first Draft in April 2003 followed by an open discussion/ consultation session in June 2003. The unedited draft of the NEP was placed on the International Development Evaluation Association (IDEA) website which resulted in sharing the contribution to policy formulation at an early stage and receiving responses from interested organizations. On receiving comments from various sectors and SLEvA members, the draft was revised and presented to the Government in late 2003. It was further revised adjusting to policy changes and finally presented to the Secretary, Ministry of Plan Implementation in June 2006. Despite the time taken to develop the draft and its subsequent adoption, by the Ministry the acknowledgment by the Government that an evaluation policy is required marks the beginning of the process of institutionalizing an evaluation culture in the country. However, the Ministry of Plan Implementation has already taken meaningful steps to institutionalize the concept of evaluation more widely across government.

Implementation of the NEP is the responsibility of all Ministries and Agencies engaged in national development functions. The MPI has to provide necessary assistance and guidance, training and refresher programs to enable efficient and effective implementation of the National Evaluation Policy. The National Operations Room (NOR) of the MPI plays the role of the focal point for implementing the NEP. The Secretaries of Line Ministries are responsible for the implementation of the evaluation process in their respective Ministries, organizations or projects. (Details of the Draft National Evaluation Policy can be obtained from SLEvA Website: http://www.nsf.ac.lk/sleva/pdf/nepdraft.pdf).

Towards capacity building

As explained before, capacity building at different levels was considered as a key intervention to develop an evaluation and results based management culture. SLEvA's contributions are as follows:

- at District[1] level, together with the MPI, SLEvA conducted a series of advocacy workshops related to the MDGS and Results Based Management;
- at the central level, together with UNICEF and MPI, SLEvA conducted **MDG advocacy workshop** with underlying RBM framework **for key policy makers** in the government;
- at individual level, SLEvA conducted several *capacity building workshops on M&E (one*

---

1   The administrative unit in the country.

week duration) for government, non-government and private sector personnel; and

• again at individual level, SLEvA conducted *evaluation methodology training workshops* as fore-runner to the International Conference held in March 2007. These workshops were conducted by reputed international evaluators and was attended by all professionals from all sectors.

Towards information dissemination

Information dissemination and knowledge management were also used by SLEvA to promote an evaluation culture. Highlights of such achievements are as follows:

• National Conferences were organized with the focus on *Evaluation Practices and Development in the Public Sector, Universities, NGOs, Private Sector and Donors (2003)* and on the theme *Development Evaluation improving outcome (2004).* These were organized together with the annual general meetings of the membership to allow new methodologies and areas in evaluation to be presented as means of experience knowledge and information exchange.

• Two International Conferences were held on the themes, *Evaluation, Good Governance and Development (2001)* and *Evaluation an Essential Element in the development process (2007).* Participants from overseas and Sri Lanka had the opportunity to exchange their experiences and knowledge and learn new methodologies in evaluation and standards in evaluation.

• *Periodic meetings* were held for the members to enable interactions with visiting evaluators.

• A *newsletter* was produced for the membership with information of SLEvA activities and evaluation information including standards.

Towards networking

SLEvA has established close collaborations with UNICEF, UNDP, German Agency for Technical Cooperation (GTZ), JBIC and the American Red Cross. These relationships have helped the association in securing assistance to promote participation in capacity building workshops and conferences overseas and to conduct capacity building workshops and conferences in the country.

In the year 2006, SLEvA became a member of the International Organization for Cooperation in Evaluation (IOCE). SLEvA hosted the formation of the Evaluation Network of South Asia (ENSA) an initiative of the UNICEF Regional Office. SLEvA expects to continue to play an active role in ENSA by promoting exchange of experiences and sharing of information among member countries.

Other Government activities that have promoted the development of an evaluation culture

The establishment of the **Information Communication Technology Agency (ICTA)**, the Government body responsible for implementing the e-Sri Lanka Development Initiatives, is yet another outstanding contribution towards institutionalizing the process of evaluation in the public sector. ICTA functions under the President which is an indication of the importance placed on promoting evaluation as a discipline. It is noted that during the bidding process for selection of consultants for the Agency, the Draft National Evaluation Policy was used as reference.

The ICTA has helped the Department of Foreign Aid and Budget Monitoring of the Ministry of Plan Implementation comprehensive *web-based National e-Project Monitoring System*. The aim of the M&E system of ICTA is a move beyond measuring inputs and outputs to capture outcomes and development results. In line with the National Evaluation Policy, the monitoring system captures implementation progress as well as results of all key development projects and program and provides policy makers and senior officials with on-line and real time access to progress information. The system generates early warnings and assists in trouble shooting of problem projects and projects behind schedule. It is under pilot implementation at the moment.

ICTA decided to establish an Advisory Board to the ICTA M&E unit and SLEvA has been invited to be a member of that Board. This provides SLEvA further opportunities to contribute further towards promoting an evaluation culture and contributing towards policy formulation at the National level.

Challenges and way forward

In his address to the International Conference organized by SLEvA in collaboration with the Ministry of Plan Implementation (March 2007), the Secretary to His Excellency the President of Sri Lanka requested SLEvA to deliver the following to improve the evaluation capacity of the Government:

- assist the Government to set up small units for evaluation in the departments and other institutions capable of in-house evaluation of all development and administrative work they perform
- work with the Sri Lanka Institute of Development Administration to establish a useful, dynamic and realistic performance evaluation and appraisal scheme for public service taking into consideration the different levels at which the government functions
- market the concepts of Program Evaluation, Project Evaluation, and Personnel Evaluation effectively to the highest levels of the government.

The Secretary to His Excellency emphasized that his requests were based on Government policy, similar to the original task of developing the NEP. SLEvA considers this request as a reflection of recognition on the part of the Government of the role played by SLEvA as an independent professional body and is currently considering ways and means of meeting the request in addition to its planned activities.

Also in the paper presented to the International Conference the Director General, Department of the Foreign Aid and Budget Monitoring in the Ministry of Plan Implementation set out other planned activities for monitoring and evaluation:

- *Monitoring Performance of Ministries and Agencies:* Results Based Monitoring and Evaluation System used with technical support of the UNDP complemented with performance agreements to ensure Ministries and Institutions to manage priorities in a results oriented manner accountable for the targets they set.
- *On-going, Ex-post and Impact Evaluation:* the Department of Foreign Aid and Budget Monitoring of the Ministry Plan Implementation will undertake on-going, ex-post and impact evaluation on selected mega Projects and disseminate evaluation findings to concerned stakeholders.
- *Evaluation Information System:* for evaluation and feedback to be followed up with the establishment of a web-based Post-Evaluation Information System to ensure effective dissemination of evaluation findings.
- *Poverty Monitoring and Tracking of Millennium Development Goals (MDGs) and National Development Goals (NDGs)* also known as Mahinda Chinthana Goals (MCGs).

Much of the efforts of SLEvA have been focused at the Government and policy level. However it also recognizes the significant role that evaluation can play at the grassroots level/ village level. Consequently SLEvA intends to help beneficiaries of development interventions

become aware of the need for self-evaluation of the delivery system. This involves the development of concepts, methodologies and capacities for facilitation at the level of grass-root level development activists. This will also require advocacy of the concept and methodologies to the Government and the Non-Governmental Institutions in order to gain acceptance for the concepts – which may be easily said than done.

However, it is encouraging to note the commitment of the Government and the Non-Government sector, private sector and donor community in promoting evaluation as a culture and a discipline as expressed by their appreciation of the efforts made by SLEvA at the International Conference held in March 2007. Accordingly SLEvA looks forward to a future of success in its endeavors to promote evaluation as an integral element in the development process.

A discussion between the editors and the authors

Q. **The SLEvA strategy has been so far very "Top Down". At the end of the article you mentioned the need now to work at the community level. Why did you choose to do things in that order?**

A. SLEvA had no choice. Development activities of the Government and the NGOs and INGOs are top down and do not have a monitoring or evaluation process built into their plans other than monitoring of physical and financial progress followed by post project evaluations to meet Donor requirements. Therefore, the priority for SLEvA was the advocacy and introduction of evaluation as an integral component to the development process, responding to the prevailing scenario.

Further, SLEvA strategy has been to focus on creating an enabling environment, both policy wise and institutionally, which is considered a pre-requisite for operationalizing an evaluation culture at the various levels in the development process, primarily at community level. Therefore, it is timely to focus on ways and means of how community participation in planning, implementing, monitoring and evaluating the development interventions could be strengthened.

It is not that meaningful or feasible to introduce grass-root level monitoring and evaluation with in the context of a "Top Down" development process.

Q. **What lessons have you learned so far from SLEvA activity? What advice would you give other South Asian countries that are establishing a national evaluation body?**

A. • SLEvA has evolved in the Sri Lankan context attempting to meet the prevailing Sri Lankan development environment. Any attempt in establishing a national evaluation body should be placed in the country specific context.
- SLEvA has worked closely with the Government with a view to advocating the adoption of evaluation as an integral component in the development process and has achieved a fair degree of success and acceptance.
- SLEvA recognizes the fact that civil society organizations cannot work in isolation, particularly in promoting an evaluation culture. Hence collaboration with the Government is considered crucial as joint efforts are found to be more productive and complementary.
- SLEvA upholds voluntarism among its membership which had been a strong contributory factor. However, the limitations should be taken into consideration for sustenance of voluntarism.
- SLEvA has provided a catalytic role to professionals interested and committed to promote evaluation.
- SLEvA activities have been conditioned by limitations of resources in the form of time available to members to devote on SLEvA activities as well as financial resources.

- Enlisting support from Donor Agencies for promoting an evaluation culture and sharing of experience as the resource based of a voluntary organization is limited.
- Capacity building and networking contribute to the success of promoting an evaluation culture.

Q. So from those lessons, what are the must dos and must not dos?

Some must dos

A.
- Uphold the voluntary nature of the organization.
- Focus on advocacy and promotion of evaluation as a discipline.
- Collaborate with the Government for policy influence for the cause
- Accept a complementary role
- Network with other development organizations for sharing of experience
- Uphold the independent and apolitical nature of the organization
- Focus on capacity building at all levels.

Some must not dos
- Should not become a commercial enterprise
- Should not have political biases.

This article started out life as a description of the Pakistan Evaluation Network. But as the subsequent dialogue between the editors and the author developed, something more personal, more reflective and richer started to emerge. What we now have is an open and honest description of the challenges faced by those wishing to promote useful evaluation within program management contexts that, historically, have been anything but evaluative. And then some insightful ideas of how these challenges can, and are, being addressed.

Evaluation challenges in Pakistan and establishment of Pakistan Evaluation Network – PEN

Khadija Khan, Ph.D

Introduction

The objective of this paper is to describe the evaluation scenario in Pakistan, identify issues and draw some conclusions for future action. I begin by briefly defining some commonly used terms and clarifying their relevance to the development process. This will provide a rich picture of the wider context for readers of this journal. Since extensive literature is available on M&E, I do not get into detailing various concepts, models, standard practices, tools and techniques.

Brief Definition of M&E

The concept of evaluation is better described together with monitoring in the perspective of commonly known phrase of M&E i,e. monitoring and evaluation. 'Monitoring' is a practice through which programme/project progress is tracked, problems are identified and corrections are made, usually with shorter turn around time so as to keep the work going in the right direction. 'Evaluation', on the other hand is a systematic assessment of various aspects of the programme/project to reach some conclusions about the performance as well as success or failure of the programme/project. Therefore assessing the effectiveness of policies driving a particular programme/project is central to evaluation. Evaluation gets its input partly from regular monitoring and partly from specific exercise at mid-term or end-term evaluation of the project. Additionally, there are also 'Impact Assessment' studies undertaken to gauge changes that occur in peoples' lives due to the interventions of the programme/project.

Evaluation Scenario in Pakistan

Pakistan reached its 60th birthday on 14 August 2007. Being a developing country, it receives significant financial and technical assistance from International Development / Financial Institutions (DFIs/IFIs), bi-lateral and multi-lateral donors, consortiums and UN Agencies in the form of soft loans and grants for development. The bulk of the development funding is channelled into the Public Sector Development Program (PSDP) through the government system. To better understand the evaluation scenario in Pakistan, it would be useful to provide some background information on current planning process.

Conventional Project Planning Cycle

All professionals concerned with planning in Pakistan, including those working in regular ministerial/departmental (non-development) or development planning areas are familiar with 'Project Planning Cycle' and 'Annual Work Plan' as well as the longer term work plans such as the 5-Year or 10-Year Work Plans prepared at the country, provincial and ministerial level. The conventional Project Planning Cycle has five steps:

    i.   Feasibility Study;
    ii. Project Plan;
    iii. Implementation (includes monitoring & mid-term evaluation);
    iv. Completion; and
    v. Evaluation.

These steps of the Project Cycle are called PC-I, II, III, IV and V.

While there is a focus on results, the emphasis is on the intended results rather than the actual results. The government department responsible for planning such as the Planning Commission stays at the helm of affairs and consolidates all the individual plans into a national document and undertakes the process for approval and allocation of resources with input from two relevant authorities - the Central Development Working Party (CDWP) and Executive Committee of National Economic Council (ECNEC).

Normally, projects are taken from PC I through PC III stage i.e. up to the 'Implementation' stage during which monitoring as well as random inspections are conducted and reports are prepared. Only those projects where mainstreaming is required are taken up to PC IV stage as the 'Completion Report' is obligatory to shift a development project (with specific development funding) to regular budget. It is our understanding that very few projects are taken to PC V stage.

Development Project Planning Approach

An alternative to what is currently happening in Pakistan would be to embed M&E from the first step and run right through the project cycle. This would mean that pre-feasibility study and needs assessment to diagnose problems and ascertain needs would be done through participatory processes that genuinely capture the needs of the community. Traditional top down planning alone would not suffice. The Key Performance Indicators (KPIs) related to inputs, outputs, outcomes and impact would be defined together with stakeholders/communities and an M&E mechanism evolved jointly to provide relevant quantitative and qualitative information periodically using various sources at administrative and community levels. Similarly periodic/final evaluations and impact assessments would be conducted with stakeholders' participation. Mid-course corrections can be made as necessary. Lessons learnt from evaluations and impact assessments would provide guidance for future planning and therefore used to evolve best practices/standards. When widely shared, the experiences turn into common basis for knowledge creation.

The Advantages of the Alternative Approach

The advantages of using the alternative approach rather than the conventional approach is that the development programmes are well monitored and evaluated. It would provide a valuable opportunity for learning lessons and making use of the experiences and knowledge gained at the individual and institutional level.

The alternative approach is derived from ground experience. A small group of NGOs are using these kinds of approaches – largely because of their tendency to involve stakeholders in development projects and their ability to reach out to the communities more effectively than the government ministries and departments. Unfortunately, it is difficult to compare the government and non-government sectors' achievements in the context of M&E practices. However, the growing influence of non-profit sector in policy making on cross cutting developmental issues such as gender equity, environment, social protection, human rights and child labour, is a result of their learning from experiences in the field that comes through continued evaluation practice.

Main Issues of Development Monitoring and Evaluation in Pakistan

There are several reasons for the Government's continued use of the conventional system.. These include issues related to :
- Conceptual clarity
- Institutional framework and mechanisms
- Capacity
- Financial resources
- Attitude towards evaluation practice
- Commitment

Conceptual Clarity

In the development context in Pakistan, the thrust of the work is still focused towards achieving objectives rather than results and on implementation rather than impact. Therefore, evaluation and monitoring are focused on measuring outputs rather than on generating feedback on outcome and impact. This practice is a result of low levels of awareness and understanding of the role monitoring and evaluation can play in supporting development processes.

There are also very few examples of published reference and training material on the theme of development monitoring and evaluation in the local context for use by development planners, practitioners and young scholars.[1]

Institutional Framework and Mechanisms

In Pakistan, there are two main official institutions that monitor and evaluate the socio-economic development process - the Poverty Monitoring Cell and the National Accountability Bureau. While the former has a specific focus on implementation of Poverty Reduction Strategy Programme (PRSP), the latter focuses on investigation of corruption and accountability at the national level. In addition, there are some research institutions that primarily collect statistical data such as the Federal Bureau of Statistics, Population Census of Pakistan, Pakistan Institute of Development Economists, Agricultural Research Council and some sector based research and training institutes. The institutional arrangements cover only the government sector. The non-profit sector works out its own

---

1   Mostly the World Bank, ADB, International and UN agencies have developed and placed material on their websites for public use. Recently NGORC – NGO Resource Centre, Karachi, Pakistan, has dedicated the quarterly journal for July-October 2005 to the theme of Monitoring and Evaluation.

mechanism organization by organization. There is no centralized authority or entity to provide guidance, set standards and undertake the M&E activity on behalf of the sector.

Since, there is no institutional mechanism across the board, there is an overlapping of functions between project planners/supervisors and implementers with regard to monitoring and evaluation.

A progressive development is that the Planning Commission of Pakistan has developed a Project Monitoring and Evaluation System and tested it with the assistance of the Asia Development Bank (ADB) and also conducted a number of ex-post evaluations. This system could be used to more effectively facilitate decision making by developing and making available a database of monitored projects.

Systematic monitoring and evaluation of the development sector has also started to get attention of the Planning Commission. There is a recognition that the planning process needs efficient management tools such as an authentic database and periodic evaluation findings/lessons learnt to be able to make quality decisions. This is a good starting point.

Capacity

As in many other South Asian countries, there is a lack of professional expertise and skills to conduct effective monitoring and evaluation. Development programs have become so diverse that for every new theme, specialized knowledge and expertise is required and this can be challenging.

International organisations such as UN agencies and financial institutions have realized the problem and are trying to strengthen the Government's capabilities. However, this tends to be only in their specific area of work and thematic interest. For instance, the ADB[2] is assisting the government in enhancing capacity for project monitoring and evaluation including the financial management within the framework of its financial assistance to Pakistan. The United Nations Development Program (UNDP) is playing the major role in the PRSP (Poverty Reduction Strategy Program), to design and carry out a comprehensive M&E system that includes critical mechanisms such as performance monitoring and impact assessment. UK Department for International Development is also supporting the Planning Commission of the Government of Pakistan in monitoring poverty and establishing a poverty line.[3]

Financial Resources

Normally due to chronic resource crunch, authorities try to economize on the project's administrative costs and therefore do not carry out regular monitoring and evaluation. Except under the coverage of capacity building as described above, there is no large funding for specific investment in strengthening the area of development monitoring and evaluation at the national level. An exception is the Earthquake Reconstruction and Rehabilitation Authority (ERRA) who have established an M&E system to ensure that the investment is managed efficiently.

2   ADB approved a TA on Strengthening Portfolio Performance and Monitoring in November 2002, while a TA on Strengthening the Financial Management Capacity of Project Management Units is earmarked for 2003.
3   Government of Pakistan Paper on Planning Commission's Activities.

Attitude towards Development M&E

Then there is the issue of attitude that determines the need and utility of M&E. Even if it is included in the agenda, the actual implementation is not geared to find results and draw lessons for future application. There are several reasons for this. One relates to the consequences of an evaluation. At times, monitoring and evaluation have been used for punitive measures rather than for enhancing performance through efficiency, efficacy and transparency. This creates fear and prevents most public officers from supporting monitoring and evaluation.

Evaluation carries a further risk of inappropriate accountability. Many a programs fail for reasons that are out of the influence or control of program managers (eg lack of pre planning, needs assessments and proper technical and financial assessments), Yet, program managers bear the brunt of the criticism, losing their jobs and professional reputation. On the other hand when projects succeed everyone claims credit.

As far as the operational matters are concerned, the debate is on as to who should conduct and lead monitoring and evaluation and who owns the results.

Committment

Political will to adopt effective monitoring and evaluation could be strengthened by creating right conditions (eg institutional mechanisms, financial resources, increased awareness). There is also increasing external pressure. The World Bank report on Pakistan Public Expenditure Review (October 1988) recommended strengthening monitoring and evaluation practice in order to ensure appropriate allocation and management of public sector resources.[4]

Strategy to address the Issues

A strategy has been developed to address these issues by the Pakistan Evaluation Network (PEN – see later). The main aspects of the proposed strategy are:
- Creating awareness, disseminating information, building knowledge and enhancing skills across social, economic and political sectors;
- Developing adequate institutional framework, systems, mechanisms, tools and techniques for monitoring and evaluation of government and non-government sectors;
- Networking within the domestic institutional framework and outside with international organizations; and
- Encouraging talent for innovation, experimentation, testing and exchange of new ideas through collaboration.

Creating Awareness, disseminating information, building knowledge and enhancing skills

The idea of clubbing together four important themes under one key aspect is meant to emphasize the need for interlinking and complementing each other to maximize the impact on the recipients and users of awareness, information, knowledge and skills.

---

4   Pakistan Public Expenditure Review by the World Bank - October 1998

The table below shows how these four interlinking factors can be implemented[5]

Concept of the Awareness Framework – Interlinking Themes

	What?	Creating Awareness	Disseminating Information	Building Knowledge	Enhancing Skills
Creating Awareness	How?		Conducting discussions and meetings; writing and printing ideas, information, facts and figures	Using discussions notes, written and printed ideas and information, documenting and publication	Learning from participation and verbal and written sources of information
Disseminating Information	Why?	Making relevant information available to stakeholders enabling them to participate in the decisions making process that affects their lives		Adding information to existing knowledge for creating new knowledge through analysis, evaluation and reporting for dissemination	Providing input to skills enhancement, including diverse options available for people desirous of learning and improving their skills
Building Knowledge	When? and/or Where?	On receiving input from awareness activities, making solid and concrete ideas available for people across board	On receiving information, consolidating information and documenting ideas to establish sound knowledge based on technical, thematic and management issues for further dissemination		On arising training need, making training and reference material available to people for skills enhancement
Enhancing Skills	Who? and (for) Whom?	All stakeholders with relevant information, knowledge and skills conducting trainings, discussions, exchanging experiences, collaborating in documenting ideas and publishing books that build the overall capabilities and capacities in the development monitoring and evaluation			

Model 1 evolved by Khadija Khan 2006 - Presentation made at the 2nd International Conference of Malaysian Evaluation Society – MES on 27-31 March 2006, Kuala Lumpur, Malaysia

5   It is to be read as,' How is Creating Awareness linked to Disseminating Information?; or Why is (what for) Disseminating Information linked to Building Knowledge? and so on.

Developing Adequate Institutional Framework and Systems

The second aspect of the strategy focuses on developing an adequate institutional framework and systems. The following three steps are proposed to create and sustain a multifaceted institutional monitoring and evaluation framework :

- Gear up the operation of existing institutions;
- Enhance current institutional capacity by using the above awareness framework to address emerging challenges;
- Build mechanisms for sustaining the institution over a long period of time.

The first two items will be addressed first by a survey to assess capacity of current systems, mechanisms, tools and expertise. This analysis will establish the need and focus for building new institutions with specialized knowledge and skills at various levels in order to undertake regular monitoring and evaluation of both, on-going and future development initiatives. The final task will be to mobilize resources and create an enabling environment to sustain the institutional framework.

Networking within National Institutions and with International Organizations

As the institutional framework gears up, networking among national institutions will develop. It will lead to further extension of the network to international organizations for collaboration on addressing not only the above challenges, but also any future emergencies.

The international networking could be multi-faceted and multi-layered. It would also encourage grass root level institutions to come together for learning from each other and sharing their local knowledge with institutions in other countries. Following that will be the networking of networks such as Pakistan Evaluation Network (PEN) collaborating with Malaysian Evaluation Society (MES) and both meeting with International Development Evaluation Society (IDEAS). This has already started with the establishment in 2007 of the Evaluation Network of South Asia (ENSA), a regional network of five evaluation societies

Encourage Talent for Innovation, Experimentation, Testing and Exchange of New Ideas through Collaboration

Finally, in order to make the strategy work in the ever changing development scenario, there needs to be incentives to develop innovative methodologies, tools and techniques on monitoring and evaluation. Platforms for launching new ideas and collaboration opportunities for sharing experiences could be provided for motivated professionals and young scholars.

Proposed Implementation Plan

For the implementation of the above strategy, a plan is being proposed with the following main activities:

- Vetting strategy by stakeholders;
- Mobilizing funds for implementation of final plan;
- Preparing kits with informative material for awareness;
- Conducting systematic, multi-level training;
- Developing M&E designs for selected programmes/projects; and
- Re-assessing capacity building needs

The plan could be implemented in phases so that its effectiveness could be assessed periodically and modifications incorporated if required.

Vetting Strategy by Stakeholders

The strategy will need to be placed before a plenary of stakeholders for review. This plenary has to be tripartite i.e. government, DFIs/IFIs/donors and implementing non-government organizations including communities representatives. The above plan and models will almost certainly be modified to incorporate the needs of the stakeholders.

Mobilizing Funding

Funding is essential. Most of the donor agencies in Pakistan are already aware of the resource crunch faced by the government departments and some are assisting in specific situations to finance M&E capacity building e.g. ADB, UNICEF and DFID. Even if bulk funding is not possible to obtain from one source, part of the plan could be sponsored by various donors.

Preparing kits with informative material for awareness

The plenary would be the first step in creating awareness and identifying 'Agents of Change' who can carry the message to other audiences. In order to strengthen the learning process, adequate material based on national and international experiences has to be developed and packaged in multi-level kits.

Conducting Systematic Multi- Level Trainings

While preparing the resource material, resource organizations/individuals would be identified to deliver training. Participants would include policy makers, planners, top management, middle management, program managers, field workers and community representatives.

Developing M&E Design for Selected Projects

To make the training worthwhile, workshop methods would be used where groups will develop and implement M&E designs for their selected projects. If this worked well, the designs could be replicated in similar projects.

Conducting re-assessment of capacity

A re-assessment of capacity would be conducted at various levels during the implementation of M&E design for future activities.

Mainstreaming the M&E Design

The final step in the implementation plan would be to mainstream the M&E Design. During activity 5.5, groups will prepare M&E designs for specific projects and test in their departments. Once the design is tested and refined it can be mainstreamed in the organization.

A discussion between the editors and the author

Bob

This is an ambitious plan. Early in the article you suggest that this is best carried out by the Pakistan Evaluation Network (PEN). Tell me more about PEN.

Kahdija

Pakistan Evaluation Network (PEN) was established in 2004 by a group of development sector practitioners whose concern was to facilitate both the government and non-government organizations to effectively carry out monitoring and evaluation of development programs through creating awareness and building capacities.

There were 15 founder members, mostly from the development field, but this has grown to 57 members including 5 institutional members

Bob

Why was PEN necessary?

Kahdija

Pakistan has an active civil society represented by non-government/non-profit organizations, media, academic institutions, citizens groups and, of course, individuals. All these groups have in one way or another expressed their concern over the performance of development programs and their impact on communities and individuals. Their informal oversight does help the government identify areas for improvement. However, PEN adds a particular focus to this activity with its specific belief that monitoring and evaluation is essential to ensure good governance, professionalism and high performance

Bob

So what has it done so far?

Kahdija

The initial period was focused on organizational matters such as developing a proper mission statement, objectives, program and range of activities. Then, it established an executive body / board constituting the founder members to manage the affairs of the Network. Since all 15 members in the first year are actively engaged in development programs/organizations, it has been difficult for them to meet face to face. Therefore e-mail communications have became the norm among the members.

Currently PEN is undertaking a range of activities:
- Carrying out advocacy with government departments, in particular the Planning and Development Division which is responsible for preparing the national development plan and conduct monitoring and evaluation of PRSP (Poverty Reduction Strategy Program) and MDGs (Millennium Development Goals).
- Proactive participation in the development process through close interaction with the government departments, UN and international agencies and civil society organizations.
- Conducting regular training of its members and members of partner organizations on relevant themes.

- Disseminating information and knowledge through issuance of brochures, PEN newsletter and e-mail communication.
- Making efforts to mobilize funds for enhancing the scope of activities to benefit its members and members of partner organizations.

Bob
In your paper you wrote about networks of networks. Who does PEN network with?

Kahdija
PEN has strong network with national organizations such as government departments, private sector and non-profit social sector organizations including trade unions across Pakistan.

PEN is also well connected internationally. It is an Associate Member of International Organization for Cooperation on Evaluation - IOCE, USA. It also has the intellectual and technical support on the subject from international forums i.e.:

- International Organization for Cooperation in Evaluation - IOCE, California, USA
- International Development Evaluation Association – IDEAS, South Africa
- Program Planning & Evaluation, Inc. Michigan, USA
- Malaysian Evaluation Society, Kuala Lumpur, Malaysia

It was also a founder member of ENSA – the Evaluation Network of South Asia

Bob
You clearly see PEN taking the lead role in promoting the M&E Strategy and Plan. That's a big task – what do you see as the main challenges?

Kahdija
PEN faced the kinds of challenges described in the first half of the paper. It started from developing a common understanding among its own members regarding monitoring and evaluation and building their capacities to perform their respective jobs better to set an example. Although a lot of moral encouragement and support was available, there still remained a large resource gap that restricted PEN from enlarging its scope of work. Time was and still is another rare commodity among professionals who have to rely mostly on e-communication for information and knowledge sharing.   So the main challenges are time and resources.

The relationship between program monitoring and program evaluation is much debated. Some people regard them as synonymous, some see them as separate but linked and some regard them not only as separate ideas, but often opposed to each other.

These debates can seem separated from reality. You've got to start somewhere, and to an extent it doesn't really matter where you start as long as you keep going and finish. This article describes just that kind of strategy – first try and get some order in to the situation, start to exploit the potential, reflect on what's been achieved and move to the next stage.

Instituting a Standardized Monitoring and Evaluation System for Assessing Development Performance: An Experience from Bhutan

Karma Galleg
Alexandru Nartea

This paper describes the experience of developing and instituting a standardized monitoring and evaluation system in Bhutan. The paper is divided into four sections. The first section provides the background outlining the rationale for developing the standardized system; section two covers briefly the objectives and key components of the National Monitoring and Evaluation System (NMES); the third section discusses the main lessons learnt and the challenges in developing the system; and ends with a conclusion in section four.

Background

Since the inception of First Five Year Plan in 1961, Bhutan has pursued a holistic development approach. Based on the vision of Gross National Happiness, it encompasses economic development, preservation of cultural and traditional values, balanced and equitable development, and environmental conservation, to enhance the well being and happiness of the Bhutanese. As a result,, in the last four decades of planned development, considerable achievements have been made in every aspect of development. The economy has grown at an average rate of nearly 7% per annum for more than two decades. The life expectancy has increased to 66 years, infant mortality rate has reduced to 40.1, and primary school net enrollment ratio has increased to 79.4%.

Despite these achievements, difficulties have been encountered in identifying the relationship between the development interventions and the results achieved. A key challenge has been to assess the impacts of development plans and in particular, to understand what difference they have made in improving the lives of the people. Like most developing countries, Bhutan lacked a good monitoring and evaluation system to effectively monitor, evaluate and report on the effectiveness of its development interventions and accomplishments. The main problems faced were:

- Absence of proper M&E institutional mechanism and clear M&E responsibilities at different levels
- Different systems and associated reporting requirements and formats of government agencies and development partners
- No coordination between the different systems
- Lack of systemic use of reports and data for learning and decision making
- Deficient feedback mechanisms on the reports submitted

- M&E limited to physical and financial inputs, as opposed to focusing on higher level/longer term results, such as outcomes and impact
- Limited capacity for conducting M&E.

In the absence of appropriate M&E mechanisms, it was difficult to gauge the effectiveness and impact of development plans and programs. While concerns have been raised time and again on the lack of holistic and systemic approach to monitoring and evaluation of the Five Year Plans, it was only in 2003 that concerted effort to develop a standardized National M&E System was spearhead by the Planning Commission.

National Monitoring and Evaluation System (NMES)

The need for a standardized approach to managing M&E required an integrated system linking various institutional responsibilities and supporting the work and knowledge flow. A framework was also needed to incorporate the expected and achieved results of the national development plans.

Objectives of NMES

The overarching objective of the NMES was to institute a standardized system for monitoring and evaluation of development plans and programs in the country.

Specifically, the NMES aimed to achieve the following objectives:

1. Provide uniform and consistent M&E procedures for line ministries/agencies and districts for monitoring and evaluation of all development policies, programs and projects planned under the Five Year Plans
2. Focus greater attention in assessing achievements of outcomes and impacts of development plans and programs
3. Promote greater efficiency, transparency, and accountability in the use of national resources
4. Provide critical information to decision-makers for timely remedial actions and policy directions and planning.

While the system was standardised, it also provided high degree of flexibility to the individual program managers, units, sections to set up their own M&E mechanism for their own internal management purposes.

The system was developed through a participatory process involving all relevant stakeholders from the line ministries/agencies and districts. A series of consultative workshops and meetings were held to foster a sense of partnership and inform thinking around system development. Accessing expertise and knowledge of diverse stakeholders in this developmental stage ensured that the system was relevant and suited to their needs.

The Planning Commission has now completed the development of the conceptual framework of NMES. The National Monitoring and Evaluation Manual was published in 2006.

Main Components of NMES

There are two main components of the NMES:

- M&E institutional set-up and procedures
- a web-based computerized system called the Planning & Monitoring system (*PlaMS*).

A more detailed description of the M&E institutional set-up and procedures is available in the National M&E Manual. The manual will serve as a guide for ministries, agencies, Dzongkhags (districts) and Gewogs (blocks) in undertaking systematic monitoring and evaluation of their plans and programs.

The web-based computerized system (*PlaMS*) is a centralized data collection and management system. It enables real time online data entry, storage, analysis to generate various planning and monitoring reports on development plans and programs.

The two components are interrelated. The institutional set-up and procedures describes the planning, monitoring and evaluation process, frequency of data collection, information requirements and responsibility of the different groups involved. The *PlaMS* on the other hand is the platform for entering, analysing and reporting on this information. In addition, the computerized system allows quick and secure real time information flow and retrieval, thus strengthening inter-institutional coordination. In order to ensure consistency and efficiency in planning, budgeting and reporting systems as well as coordination between the Planning Commission, Ministry of Finance, line ministries/agencies and local public administrations, efforts are underway to integrate the *PlaMS* with budgeting and accounting systems of the Ministry of Finance.

M&E Institutional Setup
Monitoring and evaluation in the country will be conducted at four levels of functional and administrative structure:
1. National Level
2. Central Ministry/Agency Level
3. Dzongkhag (District)Level
4. Gewog (Block) Level.

At each level, a M&E Review Committee will be instituted as a review and decision-making body for issues relating to development policies, programs and projects.

A M&E Coordinator or focal point will be appointed at each level for coordinating M&E activities and supporting the M&E Review Committee.

The cross-institutional linkages between M&E Coordinators and focal points will be strengthened through establishment of M&E networks which will help pull together the existing capacities in the M&E area to sustain cooperation, knowledge sharing, mutual support, capacity development and learning based on local experiences.

Lessons Learnt to date and Challenges

One of the unique features of NMES is that it will be used as a common M&E system across all line ministries, agencies, districts and blocks in the country. The process of developing and instituting this standardised system across all functional and administrative levels has been a slow and complex process. The NMES system had to fulfil the needs of diverse stakeholders, which was often difficult as many of the ministries/agencies had or were already in the process of developing their own procedures and systems. However, unlike in other countries, given the relatively small size of the Government functionaries and strong linkages between different levels in the Government, it was highly feasible to adopt a standardized system in Bhutan.

Our reflections
In the process of developing the system, some important lessons have been learnt and challenges identified that need to be addressed for ongoing successful implementation of NMES. Some of the most critical learnings are:

1. **The need for a high degree of advocacy and championship for NMES**
   There is strong political will and support from the decision-makers in the Government to build a credible and effective M&E system in Bhutan. Nevertheless, the actual success of this effort will depend on a lead agency taking the initiative to advocate and guide stakeholders to implement the NMES. In the context of Bhutan, the Planning Commission has a critical role to play and champion the adoption and implementation of NMES in the public administration. However, this will require sufficient time, resources and dedicated staff in the Planning Commission Secretariat.

2. **Creating sustained cooperation from the stakeholders**

    As stated earlier, the purpose of the NMES is to establish a standard M&E system for all the sectors involved in the implementation of Bhutan's Five Year Plans. There is need for ongoing consultation and involvement of partners throughout the entire implementation phase in order to promote ownership of NMES by all the stakeholders.

3. **Building capacity at all levels**

    One of the biggest challenges in implementing the NMES is the lack of trained M&E personnel in most of the sectors and in the districts. Substantial efforts will need to be devoted particularly in the initial phase of implementation for building capacity to manage M&E activities and ensure sustainability of the system. The new features of the system will have to be understood and embedded into the organizational culture so that over time it becomes "*a way of doing business*" at all levels. Consequently guidelines and training need to be delivered to the users in the public administration once the system is launched.

4. **Harmonizing donor M&E requirements with NMES**

    In spite of increased donor coordination, many donor-funded programs and projects, until now, required different M&E mechanisms to fulfil the donor agency's information needs. For instance, projects initiated with cooperation and support from different donors are required to prepare separate reports for each donor. This has led to a multitude of reporting requirements for a single agency. There is also need for dialogue with donor agencies in Bhutan to encourage greater harmonization of data collection and reporting on monitoring and evaluation. NMES is flexible enough to accommodate as far as possible the specific requirements of the stakeholders including the donors.

Conclusion

The evaluative culture in the current public administration system in Bhutan is still at a nascent stage. While there is strong acknowledgement by senior policy makers about the value of using evaluation findings as a means for rational and effective decision-making, as of today, not much has been realized in practical terms. The reason has been mainly due to the absence of leadership from a lead agency in the Government. The proposed introduction of NMES by the Planning Commission is directed towards creating this leadership and building a strong evaluative culture across the system.

The NMES particularly the information management component of the system – the PlaMS is a monitoring and a compliance tool for ensuring consistency and efficiency in planning, budgeting and reporting. NMES also provides mechanism to integrate evaluation and impact assessments as monitoring data will feed into the evaluation process. However, it is expected that evaluations will continue to be promoted as independent exercises.

Since the NMES is yet to be operationalized, it is too early to assess the success of the system. The NMES is expected to be implemented from July 2008 for monitoring and evaluation of country's Tenth Five Year Plan. Nevertheless, with strong political support from the Government, it is hoped that a successful system will be in place in the future.

The ultimate goal of NMES is to promote efficiency, effectiveness and transparency in the public sector management. Can this goal be achieved? What are its impacts? Let's wait for the ex-post evaluation results from few years from now.

The editors thought it would be valuable to follow up on some of the issues raised by authors of this paper. What follows is a conversation between Alexandru Nartea and Bob Williams.

Q. Much of the focus in the paper is on monitoring processes rather than evaluation processes.

A. I would probably agree with you that NMES is more of a monitoring tool intended to help track progress towards the expected results of the Five Year Plan/Multi Year Rolling Plan/Annual Plan (which is all part of the same planning process and feed into each – the results logic model applied also in other agencies and Governments – a one year expected result contributes progressively (year by year) to the realization of a five year/higher/longer term result).

However, going back to the planning outlines and results logic model, often the inputs and activities (and probably outputs) can be measured through ongoing monitoring. As we move to higher level results such, outcomes and impacts (which are at a program/sector level) the type of indicators which are defined may only be verified through more complex means, such as administrative data, surveys, assessments etc. Could this be part of evaluation? The planning framework with its computerized replication defines and standardizes all these types of results (activity, output, outcome, impact) and its respective M&E framework (indicators, baselines, targets, MoV), plus the reporting formats and periodicity (quarterly, annually, five year). So, I would say, the system is about both of them.

Q. In terms of being a monitoring tool is it about gaining some degree of compliance or consistency?

A. Both. Through having one system we make sure that, for example, the Ministry of Health, would use the same type of results statements (impact, outcome etc.) and respective indicators just as Ministry of Agriculture and any other ministries/agencies. Also we are making sure that the definition and the actual formulation of the result is consistent across agencies. Otherwise, can you imagine how confusing it would be to evaluate and report on the progress made of the Five Year Plan, when, for example, one says objective, the other says impact, or when each agency has a different understanding of the concepts. Not to mention that we define indicators for the results statements, so we would also have different types of indicator - higher/lower. And with that when it comes to reporting, every structure would have different formats and reports. So yes, consistency.

Q. If it is about monitoring and evaluation, consistency and compliance, then how do you see the potential conflicts and contradictions between the four being handled?

A. No conflicts from my point of view. As I said, I see ongoing monitoring being used at the input, activity and probably output level, while more complex data is gathered for the outcome and impact level indicators. I often see monitoring being underestimated. Knowing that evaluations usually take place at certain intervals, we can loose the momentum for taking corrective measures if we wait till we reach the evaluation stage. The activity is usually well in progress and funds heavily utilized by the time we evaluate. Monitoring feeds in critical information in a more frequent and timely manner. Why can't the ongoing monitoring observations on the progress of implementation of a certain activity or project add pressure to undertake a deeper analysis/evaluation?

Q. Is there an evaluative culture in the current system, and if not what challenges have been (or are anticipated) in developing one?

A. The concept of "evaluative culture" may have different meanings. However, and no matter how we understand it, it is hard to say there is a well developed evaluative culture in the current system. The whole idea of the NMES is to systematize the evaluation function in

the public administration, which could fit into the concept of building an evaluation culture. This would basically mean planning for evaluation, designing and assigning responsibilities for evaluation, budgeting for evaluation and standardizing evaluation requirements and practice within the different structures. It would also create demand for evaluation since the computerized system (with its periodic reporting requirements) would basically ask for data/information on progress towards the expected results.

Although this is work in progress (so not yet in place) we can well say that the demand for information/data as part of evidence-based planning, program formulation, decision making is presently quite stron g in Bhutan. That is quite recent. As never before, increasingly, and at different levels, the demand for data is felt more and more. We may even say that the demand is greater than the supply. Again, the idea of the computerized system is to facilitate easy input and retrieval of data, which in turn would facilitate meet up to the demand.

The challenges are:

1. evaluation is part of everyone's responsibility, and not only an M&E officer, which would rather have a coordinator, support role. I would probably not be mistaken if I say that we have the same issue in UNICEF. It is about institutionalizing the function of evaluation across the board. Because at the end, what we evaluate is either a project, program, sector/plan etc. and the findings are equally useful for program implementers, sector planners and decision maker, not just the M&E officer.

2. changing the way people are doing development work. Whilst many believe that M&E is implicit to development work, the reality is that practitioners are focused more on implementation of the activities, projects, programs and plans. Evaluation comes secondary if not forgotten entirely.

Q. **What's the "theory of change" operating here? What assumptions are you making about the change process?**

A. The system will become a necessity and a way of working built into responsibilities carried out by all: whether they are program people, M&E officers, decision makers, managers etc, or budget/accountant officers. If we succeed with the integration of planning and budgeting systems (the work done presently) we will add the budgeting component which will make the system even more indispensable.

It will take great efforts to make the system fully functional. We are embarking on a big project. It is one thing to have it integrated and technically ready, and quite another to have it embedded and utilized by the entire public administration. Training and resources therefore will be needed. And we established an ambitious target to have this implemented as much as possible by the beginning of the 10th Five Year Plan, which is mid 2008.

Q. **What evaluation lessons were learned developing this framework?**

A. Interesting question: Karma and I realized how important the knowledge and skills of Results Based Management are in building and refining this system. And since the computerized system is a replication of the planning process in Bhutan we basically ensured we had an impact on key documents. This is because at the end of the day much starts with planning, and planning is about formulation of expected results and indicators to measure those results. The computerized system, being also a replication of the planning process with its documents and respective M&E framework, goes hand in hand with the RBM approach to development.

Another lesson learned is that much of the work is more about how to integrate two computerized systems into one. The two systems were developed in parallel, two processes not informing each other.. I would say, it is more challenging to integrate something already existing than to build something new (already integrated) from the beginning

Q. Is your learning relevant to all countries, or is it relevant to a few countries, or just Bhutan?

A. Bhutan's approach of standardized M&E could be replicated in smaller countries like Bhutan where the administrative and government functionaries are small. It could also be replicated for a program that is implemented nation-wide in a particular country.

Biographies

ANNE COGHLAN

E-mail: anadhaka@dhaka.net

Anne Coghlan is an independent consultant residing in Dhaka, Bangladesh and currently is a long-term participatory monitoring and evaluation consultant to the Government of Bangladesh and UNICEF HIV/AIDS Prevention Project (HAPP). She has nearly 15 years experience designing and implementing participatory and mixed-method M&E systems, particularly with community-based HIV/AIDS programs in Africa and South Asia. Anne also teaches program evaluation at BRAC University in the School of Public Health and the Development Studies Program, and is a founding member of the Bangladesh Monitoring and Evaluation Network (BDMEN). Anne received her PhD in Program Planning and Evaluation in the Department of Human Service Studies, and a Masters in City and Regional Planning, both from Cornell University.

RUSSELL GLASGOW

(PhD, clinical psychology, Univ. of Oregon)

Russ Glasgow has conducted research on patient-centered health behavior change and self-management for the past 20 years. While at the Oregon Research Institute, he and colleagues developed and validated a number of behavioral assessment and intervention procedures for diabetes self-management, and psychosocial and social environmental factors that influence self-management. More recently, he and colleagues at the Kaiser Permanente Colorado have developed and validated instruments that assess multi-level support and resources to facilitate self-management, and that assesses the extent to which patients have received care congruent with the Chronic Care Model. Dr. Glasgow has also contributed many studies that use the RE-AIM model (www.re-aim.org).

Other recent work has focused on health technology aided interventions that work in concert with primary care settings, and offer potential for widescale dissemination. Throughout these writings, he has emphasized the importance of taking a comprehensive, yet practical, integrated systems approach to understanding, assessing and intervening to enhance consistent delivery of evidence-based, patient-centered care.

PHILIPPE GIRAULT

E-mail: PGirault@fhibkk.org

Philippe Girault is a Male Sexual Health Senior Technical Officer at Family Health International Asia Pacific Regional Office, specializing in training and education of health personnel in multi-cultural settings, training needs assessment, curriculum development, training evaluation, quantitative and qualitative research, and behavior change interventions. For more than ten years, he has collaborated with government agencies, international and national non-governmental organizations, and community based organizations in the Southeast Asia region on the development of HIV interventions targeted at vulnerable populations, particularly MSM and uniformed services, and creating needs assessment and development of male health sexual services in Cambodia, Indonesia, Vietnam, Pakistan and Thailand. He conducted several research and baseline behavioral surveys to guide HIV interventions.

GEORGE F. GROB
President, Center for Public Program Evaluation
38386 Millstone Drive
Purcellville, Virginia, USA, 20132
GeorgeFGrob@CS.com
Tel: 540-454-2888

George F. Grob is the President of the Center for Public Program Evaluation. He started this consultancy after more than 37 years in the U.S. Department of Health and Human Services HHS, many as an executive directing the work of evaluators and policy analysts. As Deputy Inspector General for Evaluation and Inspections he supervised more than 1000 evaluations, covering the gamut of health and human services programs. The studies his office produced are credited with saving billions of dollars in government expenditures and improving services in many areas, including protection of human research subjects, organ transplants and human tissue donation, child support enforcement, and nursing home care. His most recent assignment was Staff Director of the Citizens' Health Care Working Group which sent recommendations to the President and the Congress on improving the U.S. health care system.

DR. RAVI I. JAYAKARAN
SDP, MAP International
The Hurt Building
Suit 400
50 Hurt Plaza
Atlanta, GA 30303
USA
E-mail: Ravi_Jayakaran@online.com.kh
E-mail: Ravi@Jayakaran.com

Dr. Ravi Jayakaran is Chief Consultant for Strategy & Development programming with Quality Professional Interfaces (QPI-GMS®), which is a development consultancy in the GMS countries of SE Asia. He has extensive experience in integrated community development work predominantly in the Asia region, and has acquired this working in the Public Sector, the corporate sector and in International NGOs for several years, and through consulting for the ADB, UNDP, and several International NGOs in SEA for over 29 years. Participatory Wholistic development is an aspect of development that Dr Jayakaran is especially interested in.

Dr Jayakaran has written several books on integrated development programs and the use of Participatory methodology. Dr Jayakaran has just finished writing another book, entitled *Facilitating Child participation in the development process* that is shortly due for publication.

Dr. Jayakaran is based in Phnom Penh, Cambodia where his wife Vimla is a teacher at the International School of Phnom Penh.

KHADIJA KHAN PH.D

Khadija Javed Khan
Earthquake Reconstruction and Rehabilitation Authority – ERRA
Prime Minister Secretariat
Islamabad, Pakistan
Tel: +92 51 9030967
E-mail: khadija@erra.gov.pk

Khadija has been associated with socio-economic development in Pakistan for the last fifteen years with particular emphasis on policy development, strategic development, monitoring and evaluation of development programs and impact assessment to gauge change at community level.

She is currently Chief, Knowledge Management Cell with the Government of Pakistan, Earthquake Reconstruction and Rehabilitation Authority - ERRA in Islamabad.

She has written many technical papers and articles based on my experiences and also trained professionals as part of knowledge sharing and capacity building in M&E.

As a founder member of Pakistan Evaluation Network - PEN, Khadija has the responsibility to head the organization which is also acting as the Secretariat for Evaluation Network of South Asia – ENSA. She is also Secretary and Member Board of International Development Evaluation Association – IDEAS.

ALEXANDRU NARTEA

M&E Officer, UNICEF Thimphu/Bhutan

Alexandru Nartea specializes in planning, monitoring and evaluation as it relates to UNICEF programs in Bhutan. He coordinates the data collection activities and supports the planning, monitoring and evaluation capacity development initiatives for the focal points of the Royal Government. During the past year of office with UNICEF Bhutan, Alexandru has jointly with the program managers from Planning Commission and National Statistics Bureau organized several trainings on Results Based Management and DevInfo in Bhutan. He continues to support the Knowledge Management activities, as innovative way of doing development work and bringing expertise and capacity development in the country.

A young professional, Alexandru has previously done development work in Central Asia and Eastern Europe, and holds a Masters' in International Relations from Linkoping University and a Masters' in Political Science from OSCE Academy in Bishkek.

DIMITRI PRYBYLSKI

E-mail: DPrybylski@fhibkk.org

Dimitri Prybylski is a Senior Technical Officer for Surveillance and Evaluation at the Family Health International (FHI) Asia Pacific Department. He has over ten years of experience conducting applied epidemiological research, surveillance and program evaluation in developing country settings. Dr. Prybylski is also an adjunct faculty Research Associate at the University of Maryland School of Medicine. He has an in-depth understanding of epidemiological research and program evaluation methods, national biological and behavioral HIV/STI surveillance systems, health information systems, and multivariate quantitative data analysis methods. He provides technical assistance in surveillance and evaluative and operations research to country-level HIV/AIDS prevention and care & support programs.

Dr. Prybylski also provides regional guidance related to research ethics to ensure the protection of human subjects. Dr. Prybylski received his PhD in Epidemiology with a concentration in Infectious Diseases from the University of Maryland School of Medicine and an MPH in International Health and Epidemiology from the Johns Hopkins University School of Hygiene & Public Health.

MALLIKA RUKMINI SAMARANAYAKE B. A. SPECIAL (CEYLON) M.ED (COLOMBO)

Institute for Participatory Interaction in Development (IPID)
591, Havelock Road, Colombo 06, Sri Lanka
Tel: +94-011-2365521 Tel/Fax: +94-011-2587361
E-mail: ipidc@panlanka.net, info@ipidlk.org
Web Address: www.ipidlk.org

Ms. Mallika Rukmini Samaranayake B. A. Special (Ceylon) M.Ed (Colombo) presently President Sri Lanka Evaluation Association (SLEvA) and Chairperson Institute for Participatory Interaction in Development (IPID). She is professionally qualified having undergone training in a wide range of subject areas relevant to Planning, Management, Monitoring and Evaluation. She has served as a Development Specialist in Participatory Rural Appraisal (PRA), Logical Framework Approach (LFA), Participatory Monitoring and Evaluation (PM&E), Social Assessments, Poverty Assessments, SWOT Analysis, Stakeholder Analysis and Participatory Impact Evaluation etc. serving as a Freelance Consultant, Consultant Sociologist, Community Development Specialist, Facilitator and Awareness Specialist.

She has been a fellow of the Institute of Development Studies (IDS), University of Sussex, UK and has served as Participatory Development Specialist in the World Bank, Colombo Office, Joint Director, Self Help Support Program (Swiss Interco-operation) and Special Project Director, Ministry of Education, Sri Lanka, as a Teacher. She has been a Team Leader and also functioned as a member of several research teams. She is currently engaged in related work in Sri Lanka as well as abroad.

MEENAKSHI SANKAR

Martin Jenkins and Associates Ltd.
Wellington, New Zealand
Tel: +64-4-9319376
E-mail: meenakshi@martinjenkins.co.nz
Web Address: http://www.martinjenkins.co.nz

Meenakshi Sankar is a Senior Consultant with MartinJenkins, a Wellington-based consultancy firm that specializes in organizational strategy, design and change, public policy and issues management, evaluation and research, human resources management and financial and economic analysis. Meenakshi has specialist skills in policy research and evaluation, understanding of social and community development and international development issues. She is an experienced evaluator and has published papers on various topics including participatory action research, collaborative case study based work and reflections on promoting use of evaluation findings. Her particular strength lies in being able to build bridges between policy, research and practice and ensuring that evaluation generates information that is useful.

Prior to moving to New Zealand, Meenakshi was a partner in a market and social research firm in India and was responsible for expanding the social research focus of the company. She has a Masters Degree in Social Work from the University of Mumbai.

MARIANA STIRBU

Social Policy Project Officer, UNICEF Regional Office for South Asia (ROSA)
PO Box 5815, Lekhnath Marg, Kathmandu
Tel: 977 1 4419471 ext 272
Mobile: 977 98031 95623
E-mail: mstirbu@unicef.org

Mariana Stirbu holds the position of Social Policy Project Officer at the UNICEF Regional Office for South Asia (ROSA), where she supports the work of the Social Policy Cluster in public policy analysis, especially regarding the impact of policies on the situation of children, and cross-sectoral issues around the rights of the child and MDGs. Previously, she held consultancies with UNICEF Moldova and UNICEF ROSA, and prior to joining UNICEF, worked with NGOs and international development projects in Moldova, where she combined research and analysis with management of development projects.

Ms Stirbu holds an MA in Development Studies, with a specialisation in Public Policy and Management, from the Institute of Social Studies, The Netherlands, where she conducted graduate research as a Joint Japan/World Bank Graduate Scholar. Her main areas of expertise include: policy research and analysis; policy advice; program and project management; and capacity building for public and non-profit organisations.

BOB WILLIAMS
Wellington, New Zealand
E-mail: bob@bobwilliams.co.nz
Web Address: http://www.bobwilliams.co.nz

Bob is known in evaluation circles through his evaluations, his work within the Australasian Evaluation Society (AES) and the American Evaluation Association (AEA), his contributions to a range of internet discussion groups including "Evaltalk", plus a wide variety of lectures, publications, books and workshops. His speciality includes the application of systems concepts, action research and large group processes in evaluation settings.

Bob co-edited the AEA monograph *Systems Concepts in Evaluation: an expert anthology*. This publication explored the potential of systems-based approaches to evaluation. This is the first publication of its kind and can be accessed from his website (http://www.bobwilliams.co.nz). He is also a member of the Editorial Advisory Boards of the American Journal of Evaluation and New Directions in Evaluation. Bob was recently appointed to the American Evaluation International Committee.

unicef