11

The challenges ahead are at least as daunting as those already overcome: demonstrating that monitoring and evaluating findings can make a difference in the lives of people.

A Microcosm of the Global Challenges Facing the Field: Commentary on HIV/AIDS Monitoring and Evaluation

Michael Quinn Patton

Stake (2004), in a provocative and insightful article, asks "How Far Dare an Evaluator Go Toward Saving the World?" He identifies six advocacies common in evaluation studies. I want to use his framework to position this review and commentary. (The direct quotations in the six items are from Stake, 2004, p. 104.)

1. *We often care about the thing being evaluated.* Stake was one of the first evaluation theorists to clearly articulate that we do not have to pretend neutrality about the problems programs are attacking to do fair, balanced, and neutral evaluations of those programs. Who wants an uncaring evaluator who professes neutrality about homelessness, hunger, child abuse, or community violence?

I care about human immunodeficiency virus-acquired immunodeficiency syndrome (HIV/AIDS). My younger brother died of AIDS early in the epidemic. My entire family has been involved actively in AIDS Walks and other activities. The editors of this issue did not know any of this when they invited me to do this review, but it was a major reason I accepted the invitation.

2. *We, as evaluation professionals, care about evaluation.* This entire issue epitomizes that caring. Stake could have been describing this issue of *New Directions for Evaluation* when he wrote: "We care about evaluation. We want to see others care about it. We want to encourage them to do it. We promote evaluation services, our own and those of our profession.

 © Wiley Periodicals, Inc.

We favor methods that evaluate well, and encourage others to use them too. It is an advocacy we flaunt" (2004, p. 104).

A powerful verb, *flaunt*. This volume *flaunts* the potential contribution of monitoring and evaluation (M&E) to solve the global HIV/AIDS crisis.

I count myself among those who care about evaluation generally and HIV/AIDS evaluation specifically. The experiences of those involved with the development and implementation of HIV/AIDS M&E systems provide us with a macrocosm of M&E more generally. The chapters in this issue represent much of what is good about evaluation—and also much of what is wrong with it. I would ask that my criticisms be understood in the context of caring about evaluation.

3. *"We advocate rationality."* The mindset of rationality constitutes a foundation that supports the M&E initiatives and recommendations that run throughout the chapters. Logic modeling is one common manifestation of rationality in evaluation, logic being a primary tool for rational thinking and a criterion against which we evaluate rationality. It is not surprising to learn, then, that the basic M&E organizing framework endorsed by all participating agencies is based on a simple "input-activities-output-outcome-impact" framework (Chapter Two, this issue).

Harkening back to the first advocacy in this list, however—that of caring about the thing being evaluated—I wonder whether we can find a place for *heart* and emotion amid all this heady stuff. Later in this review, I will rant (perhaps irrationally) on this concern.

4. *"We care to be heard. We are troubled if our studies are not used."* The M&E approaches described and analyzed in this issue are often labeled "utilization-focused M&E." That happens to be a matter of some interest to me as well as a little knowledge, so I will opine about the extent to which the M&E systems offered here manifest utilization-focused premises and, to the extent that they do, some of the dilemmas and challenges posed.

5. *"We are distressed by underprivilege. We see gaps among privileged patrons and managers and staff and underprivileged participants and communities."* In the global context, poverty and HIV/AIDS go hand in glove. Among the privileged are evaluators. The M&E systems described in this issue cost millions of dollars and provide privileged employment to thousands. Among other things, this raises the stakes to demonstrate that such expenditures, which directly reduce funds available for treating people afflicted with HIV/AIDS, ultimately add value and benefit underprivileged participants and communities. Whenever I conduct evaluation training in developing country settings, this question is always front and center. In this issue, as in most of evaluation, this nagging question of evaluation's ultimate cost-benefit remains one of hope and idealism rather than empirical demonstration.

6. *Finally, "We are advocates of a democratic society."* Operating in a global context of greatly diverse political systems, the efforts reported in this issue supersede variations in national political systems to accomplish

the goal of a universal and standardized HIV/AIDS M&E system around the world.

There is even evidence that dictators can be information users. De Lay and Manda open Chapter One by telling the story of how President Museveni of Uganda got his initial push to deal with HIV/AIDS from President Fidel Castro of Cuba, who called and told him that eighteen of the sixty Ugandan soldiers who had been sent for training in Cuba had tested positive for HIV. Uganda subsequently became one of the exemplar nations in turning back the tide of HIV/AIDS. Try mapping that outputs-to-outcomes-to-impacts logic model chain to explain and attribute high-impact information use.

Is HIV/AIDS M&E Different?

De Lay and Manda identify four ways in which the unique nature of HIV/AIDS may pose special challenges for M&E. First is global denial. De Lay and Manda argue that nation-states have handled HIV/AIDS differently from other infectious diseases. Prevalence numbers have been diminished, suppressed, or delayed and, when reported, explained away as being infections of foreigners. In short, it has been difficult to get governments to acknowledge the scale of the problem and to intervene in a timely manner, including supporting the collection of valid and reliable data about prevalence.

This is not, however, a problem specific to infectious diseases. Problems of child abuse, family and community violence, homelessness, mental illness, unemployment, rape, and corruption have suffered, and often still do, from underreporting, data manipulation, report suppression, and findings delay. Nor are these problems limited to developing countries.

Data used to monitor donor needs for program performance: I am afraid HIV/AIDS evaluations have no special claim here. As near as I can tell, it is a nearly universal phenomenon that the data most useful to local programs fail to provide what international donors want. Guess who wins in setting data priorities? Podems (2004) just completed a doctoral dissertation on this issue studying a youth program in South Africa. She offers a framework for attempting to improve communications between funders and local implementers using evaluation as a communications tool.

Protection of individual rights versus the public good: Tensions between the need for data and the right to privacy have emerged in a number of areas, although HIV/AIDS surely intensifies the stakes on both sides. But this tension is also manifest in trying to track child-support payments, sexual predators, genetic diseases, welfare fraud, unemployment benefits, foster care, adoptions, mental health, and family violence. One might say that there is almost an inherent tension between society's need to know to inform policy decisions and individuals' desire to protect themselves from government intrusion into their lives. Orwellian nightmares dominate this debate.

Selective application of evaluation research in support of ideologies and values: Selective use of evaluation data to support political bias, also worrisome to De Lay and Manda, and appropriately so, may actually be less in HIV/AIDS than in other areas. Public health has a long and distinguished tradition of getting the data out and getting findings out in a timely way. Yes, there are inevitably short-term efforts to manipulate data for political purposes in the public health area, but the stakes are typically so high and the professionalism of health staff so well established that these manipulations and distortions find their way into the public light, especially in the age of the Internet. No, HIV/AIDS has no special privilege here. The problem is simply widespread and endemic to politics.

Tensions Between Competing Purposes for Evaluation

M&E systems are asked to satisfy a number of competing purposes—for example, accountability, program improvement, management decision making, identifying effective models for dissemination, and supporting collaboration among partners. In Chapter Two, Rugg, Carael, Boerma, and Novak suggest that the primary purpose of the HIV/AIDS M&E system should ideally be program improvement. In reality, they found that donor reporting and accountability demands burden the system and engulf local data-collection efforts to such an extent that evaluation is little used for program improvement. This is one of the most common observations in evaluation. Accountability demands routinely trump hopes for and intentions of program improvement.

We see this tension in the chapters in this issue. Chapter Two emphasizes the program improvement uses of evaluation while acknowledging a need for accountability. Chapter Five, on developing and implementing M&E "in the new era of expanded care and treatment of HIV/AIDS," emphasizes accountability, with an occasional nod to program improvement. Emphasis matters.

In a similar vein, Wilson (Chapter Six) notes in his African capacity-building analysis that most M&E frameworks are not designed as management tools. This is related to the problem of trying to have the same system serve both accountability and program improvement purposes—a good idea in theory but one that seldom works in practice. These different purposes need different data and different analysis processes.

Another universal tension is between bottom-up and top-down design and utilization processes. Anie and Larbi (Chapter Eight) describe Ghana's approach as "bottom up." It is helpful to take local people through the process of understanding and accepting indicators developed by international experts, but that is not the same as a genuinely bottom-up process in which grassroots priorities set the agenda, determine the methods and process, and develop the indicators of choice. The language matters here as well. Calling something "bottom up" when it is not experienced that way locally can lead

to later alienation and lack of use. One of the most common complaints I hear from people at the grassroots level is that they were told that a process was bottom-up when, in fact, the real agenda was acquiescence to a top-down design. Bringing the needs of national and international users together with the needs of local and regional users remains one of the great challenges of M&E. National stakeholders and their information needs typically trump local stakeholders and their information needs. It is not clear to me that that did not happen in Ghana, but I could be wrong. The point is that no evidence is offered that the process *was actually experienced* as bottom up.

To determine how use actually occurs at these different levels, we need more rigorous follow-up than is available here. Take the case of Thailand's experience, reported by Kanshana and others (Chapter Seven). To the authors' credit, substantial attention is given to M&E use. But what is offered falls short of telling readers how the results are actually used. For example:

> One of the important uses of this monitoring system is to provide data to indicate programmatic and geographical areas where coverage is not optimal so that national or local policies or practices can be adjusted to improve the program. To this end, program staff were trained on interpreting the nine standard reports; a series of regional and national meetings were held to bring together program staff and policymakers to review data, identify areas of successful and unsuccessful implementation, and identify solutions to improving the program.

So far, so good, but we are not told what actual program improvements were made, if any. The process sounds excellent, and the intentions are exquisitely utilization focused, but the unanswered question is whether anything important actually changed as a result of these meetings. Were the findings discussed actually used to make identifiably significant improvements to the program? A truly utilization-focused M&E system will include closing the loop by evaluating actual use and learning what factors enhanced use and what factors may have inhibited use, then using those learnings to further enhance use (Patton, 1997).

Humanizing M&E

I promised to return to advocacy of rationality and logic in evaluation, the third item on Stake's list. I wondered earlier whether we could find a place for heart and emotion in evaluation. Certainly I would not be so rash as to suggest a prominent place, but a place nonetheless. I was stimulated to ponder this question as a result of the overall impression I took from reading the contributions in this issue.

These chapters portray herculean efforts to create, organize, and implement rigorous and standardized M&E throughout the world. It is an

ambitious, even staggering, undertaking. To get multiple and competing international agencies and jealous-of-their-autonomy governments around the world to agree to the "three ones" principle is an impressive achievement. One can only hope that this achievement lives up to its advance billing of better and more useful data. The danger and potential downside of perfectly coordinated and centralized data collection and reporting is that it can be harder to detect corruption and correct abuses. Monopolies require careful and transparent scrutiny and regulation. The three ones principle celebrates and institutionalizes an HIV/AIDS M&E monopoly. That may well yield important benefits of comparative data, efficiencies in data collection and analysis, and global economies of scale. It may also collapse under its own weight, which is my reading of the history of many such systems.

But I digress. The central point I want to make in this section is my overall impression of these HIV/AIDS M&E systems as highlighting numbers, reports, comparisons, collaborations, indicators—and systems. All of these things are important, indeed critical. But where are the portrayals of people living with HIV and dying of AIDS? I know, I know. That was not the purpose of this issue. But perhaps it should have been *a* purpose. Perhaps it should always be a purpose in evaluation to remind policymakers, journalists, diverse stakeholders, the general public, and evaluators that there are real people behind the numbers and data reports.

If I could wave a magic evaluation wand and change M&E evaluation reporting, it would be to mandate that monitoring data always be reported with stories about real people so that the numbers do not take on an abstract life of their own. These chapters portray millions of dollars being spent to create standardized systems and coordinated systems and performance systems and tracking systems and monitoring systems and data storage systems and data-sharing systems. Where are the people? Not just the people suffering from HIV/AIDS, although their absence from these chapters is a travesty, in my judgment. But I would also welcome stories about the people at the grassroots level who enter the data into these systems and those who are supposed to use these data. I have met many of these people in the workshops I do around the world. They offer a different perspective on global, standardized M&E systems. Evaluating those systems will ultimately need to include those stories and perspectives.

A Systems Perspective

Speaking of systems and perspectives, another overall impression I came away with from reading these chapters is how deeply entrenched mechanistic linearity is in evaluation. I do not share the authors' enthusiasm for the endorsement by all participating agencies of a simple input-activities-output-outcome-impact framework. This strikes me as an especially limited framework for understanding HIV/AIDS. The two cases of dramatic national

reversal of HIV/AIDS infection rates with which I am most familiar, Uganda and Brazil, are fundamentally stories of complex, dynamic systems change. The stories differ in their specifics, but each involve multiple, overlapping interventions at different levels and in different ways to change national consciousness, community attitudes, cultural values, political priorities, and normative behaviors. Interdependent change occurred in religious communities, political policies, educational institutions, community organizing, public health, and criminal justice. Under such circumstances of massive, interconnected, and interdependent dynamic systems change across sectors, budgetary sources, and interventions (national, regional, and local), complex systems change mapping and networking models hold more promise than do traditional linear-logic models.

I had the opportunity to do evaluation training for a diverse group in South Africa. Participants included educators, health workers, police, environmentalists, park rangers, youth program staff, employment program managers, civil servants, and many others. All of them reported being involved in some way in the battle against HIV/AIDS infections regardless of their organization's mission. Firefighters, for example, were distributing HIV/AIDS information and supporting condom use. In such an environment, facing such a massive problem with such huge societal implications, the autonomous program may not be a meaningful unit of evaluation analysis. I certainly cannot imagine disentangling the various and overlapping interventions I have witnessed and heard about in various African countries. Yet, the dominant concept of evaluation in this issue is a traditional inquiry into an autonomous intervention delivered in a linear-outcomes model to isolate and attribute its impacts, including conducting quasi-experimental designs. I am openly skeptical about the utility of such designs in these circumstances. Indeed, the controls needed to even attempt such designs risk having the evaluation design interfere with the effective creation and implementation of complex and dynamic systems interventions that are too messy and emergent to be appropriately evaluated by static designs and linear, mechanistic attribution models.

Chapter Nine, on intervention research and program evaluation, epitomizes this limited perspective. Peersman and Rugg's four quality criteria for methodological rigor reflect the narrow view that, especially in dire circumstances where the stakes are high, evaluation is no different from research and must favor quantitative measures and linear, experimental attribution designs. These influential authors give token attention to alternative evaluation methods and none to complex systems dynamics.

In essence, I would argue that HIV/AIDS is best understood as a dynamic, complex systems phenomenon. To be useful and insightful, evaluation designs need to match the nature of the phenomenon. The evaluation framework celebrated in these chapters, while recommending that the simple linear approach be expanded to include the triangulation of data from multiple sources and methods, does not go far enough to embrace a

systems framework and thus fails to map the nature of the phenomenon to the evaluation approach. In so saying, and in defense of the authors, I acknowledge that the M&E approaches presented here represent mainstream evaluation thinking, which is precisely the problem.

First Do No Harm

Evaluation is not benign. In addition to potentially wasting scarce resources and diverting desperately needed resources from treatment, evaluation can do harm in other ways. One of these is evaluating what are on the face of it unrealistic and overpromised goals, adding to the pretense that the goals are meaningful. Consider the "hard targets" of the United Nations General Assembly Special Session's Declaration of Commitment being adopted in national plans, as described in Chapter Three. Take just one example: "By 2005, ensure that at least 90 percent, and by 2010 at least 95 percent, of young men and women aged 15 to 24 have access to the information, education, including peer education and youth-specific HIV education, and services necessary to develop the life skills required to reduce their vulnerability to HIV infection, in full partnership with young persons, parents, families, educators, and health care providers."

This kind of goal guarantees failure. Where do such target outcomes come from? One of the common misperceptions about utilization-focused evaluation is that the evaluator simply yields passively to whatever stakeholders suggest, acquiescing to their perspective regardless of whatever absurdities they may offer. Quite the contrary, I have argued that every evaluation is a training opportunity to increase the sophistication of those involved, especially primary intended users, and that the evaluator has a stake in the integrity and sensibility of the evaluation. When goals are patently overpromised and unrealistic, based on generic evaluation knowledge about what is likely and possible, as is often the case and certainly the case for this goal, it becomes the evaluator's responsibility to facilitate careful consideration of the implications of proceeding. I reject the perspective that our job is simply to evaluate whatever goals we are handed without commentary on their sensibility and meaningfulness. *This is where we get to be advocates for rationality*.

A Microcosm of the Global Challenges

The authors of the chapters in this issue portray, on the whole, an unusual degree of success in creating and implementing a universally accepted M&E framework, including identification and standardization of key indicators. Those involved have also established standards for quality HIV/AIDS evaluations and have shown remarkable acumen in navigating the treacherous political and cultural shoals of HIV/AIDS to establish a global foundation for M&E. In this regard, these chapters, although focused on HIV/AIDS

M&E approaches, reflect the state of the art of evaluation more generally—the field's strengths, its shortcomings, and its controversies. I have offered some additional criteria for what constitutes success that calls into question some of the achievements touted in this issue.

The successes reported here in creating worldwide HIV/AIDS M&E systems also increase the stakes in making sure that the resulting findings are useful and actually used to reduce the infection rate, reduce deaths from AIDS, and assist those living with HIV/AIDS. Given the high costs of these systems, the stakes for evaluation are high. Can evaluators deliver on the promise that better information will lead to more effective policies and programs; better allocation of scarce resources; and better outcomes at local, national, and international levels? In that regard, the authors record and document important milestones in HIV/AIDS M&E globally. But the challenges ahead are at least as daunting as those already overcome, namely, demonstrating that monitoring data and evaluation findings are worth what they cost and can make a difference in the lives of people vulnerable to and afflicted by HIV/AIDS. Stay tuned. Given the visibility and importance of the global battle against HIV/AIDS, evaluation's contributions here are likely to ripple and affect perceptions about the value of M&E systems generally.

References

Patton, M. Q. *Utilization-Focused Evaluation.* (3rd ed.) Thousand Oaks, Calif.: Sage, 1997.

Podems, D. "Nonprofit Evaluation in South Africa: A Study of Relationships Between the Donor and Nonprofit Organizations in the Developing World." Unpublished doctoral dissertation, Union Institute and University, Cincinnati, 2004.

Stake, R. "How Far Dare an Evaluator Go Toward Saving the World?" *American Journal of Evaluation,* 2004, *25*(1), 103–107.

MICHAEL QUINN PATTON is an International Consultant at the Union Institute and University in St. Paul, Minnesota.