
Forum contributions present essays, opinions, and professional judgments. Forum articles speak to and about the philosophical, ethical, and practical dilemmas of our profession. By design, the "Forum" is open to diverse views, in the hope that such diversity will enhance professional dialogue. Standard citations and reference lists should be used to acknowledge and identify earlier contributions and viewpoints. Manuscripts should typically not exceed 15 double-spaced typewritten pages in length, unless the paper is invited by the editor.

The New Environment for Development Evaluation

Robert Picciotto
King's College London

Abstract: The millennium development goals have created new challenges for development evaluation. The main unit of account has shifted to the country level. Evaluation ownership must move from donor agencies to developing countries. The recognition that rich countries have development obligations is opening up evaluation frontiers beyond aid. A transformation of evaluation priorities is needed: (a) Evaluation frameworks should give more weight to alignment with the millennium development goals, (b) impact measures of development programs should be aggregated to the country and global levels, (c) accountability should be enhanced by sharper attributions of results according to the distinctive accountabilities of development partners, (d) attribution of results to aid should be examined using methods appropriate to the situation, and (e) the asymmetry of the development evaluation agenda should be remedied by a sharper focus on the impact of rich countries' policies on global poverty reduction.

Keywords: *attribution; aggregation; accountability; development*

The development enterprise is at a crossroads. In Kofi Annan's (2002) words,

Ours is a world in which no individual and no country exist in isolation. All of us live simultaneously in our own communities and in the world at large. . . . We are all influenced by the same tides of political, social and technological change. Pollution, organized crime and the proliferation of deadly weapons likewise show little regard for the niceties of borders; they are problems without passports. (p. 30)

Thus, "glocalization" (the combined effect of globalization and localization) best defines the new operating environment for development. On one hand, in a world of extraordinary diversity, the challenges of poverty reduction are mostly local: Tailor-made development solutions

Robert Picciotto, King's College London, 52 Wynnstay Gardens, London W86UU, United Kingdom; phone: (44) 02079372404; fax: (44) 02073611660; e-mail: R.Picciotto@btopenworld.com.

American Journal of Evaluation, Vol. 28 No. 4, December 2007 509-521

DOI: 10.1177/1098214007306371

© 2007 American Evaluation Association

require “on-the-ground” knowledge and indigenous decisions. On the other hand, in today’s integrated world economy, sustainable development must overcome global obstacles and tap global opportunities; no nation can thrive unless it can compete within a global market shaped by rich countries’ policies and standards.

In this new operating environment, development evaluation will be shaped by two distinct but interrelated contexts. The first is cultural and location specific. The second has to do with evolving global structures, ideas, and interests. This article describes how these two sets of forces will influence the trajectory of development evaluation in the years to come.

The Evolving Role of Development Evaluation

Gone are the days when development evaluation largely consisted of assessing the costs and benefits of individual projects. As the ambitions of the development enterprise expanded, the demand for evaluation gradually shifted to the country level. This is because development prospects (e.g., with regard to the design of public policies or the management of public funds) mostly hinge on policy decisions taken by national governments. As democracy spreads, all citizens, including the poor and the neglected, feel entitled to bring public officials to account. This explains the rapid spread of national evaluation associations throughout the developing world.

But development impacts are also the result of policy decisions taken at the global level in the interconnected realms of international aid, trade, and finance. How poor countries’ interests are represented in the corridors of international power hinges in part on the governance structures of international organizations in which evaluations plays a part. Thus, development evaluation has become more international and more indigenous all at once, and its scope now extends well beyond aid. Trade, investment, migration policies, and the other “rules of the game” that govern the international economy have a major impact on the poor of the world.

Evaluation is an institution, and all institutions are path dependent so that rapid institutional change is uncommon. But a drastic reconsideration of the principles and practices of development evaluation has become imperative given the historic development compact forged by all United Nations members at the turn of the century. Specifically, the Millennium Declaration endorsed by all heads of state in New York in December 2000 defined development as a global partnership for poverty reduction. Two years later, the United Nations Financing for Development Conference in Monterrey, Mexico, endorsed specific millennium development goals (MDGs; Table 1) along with a battery of performance indicators.

The international agreement to achieve the MDGs was the culmination of decades of efforts by the world body. The MDGs enjoy broad-based legitimacy because they are grounded in the deliberations of international conferences sponsored by the United Nations and reflect universal human aspirations. They have put to rest fruitless debates about the meaning and purpose of development cooperation. In parallel, they have ushered in a new era for development evaluation by providing commonly agreed benchmarks for the development cooperation enterprise. If the solemn endorsement of the MDGs by all United Nations members constitutes a watershed in development cooperation, it is not only because measurable development objectives have been specified but also because distinct responsibilities and reciprocal obligations have for the first time been acknowledged by rich and poor countries alike.

First, poor countries have agreed to take charge of poverty-reduction programs, improve their governance, and reform their policies. In turn, rich countries have agreed to contribute

Table 1
The Millennium Development Goals

1. Eradicate extreme poverty and hunger	Halve the proportion of people with incomes of less than \$1 a day between 1990 and 2015 Halve the proportion of people who suffer from hunger between 1990 and 2015
2. Achieve universal primary education	Ensure that boys and girls everywhere complete primary schooling by 2015
3. Promote gender equality and empower women	Eliminate gender disparity at all levels of education by 2015
4. Reduce child mortality	Reduce by two thirds the younger than 5 mortality rate between 1990 and 2015
5. Improve maternal health	Reduce by three fourths the maternal mortality ratio
6. Combat HIV/AIDS, malaria, and other diseases	Halt and reverse the spread of HIV/AIDS, malaria, and other major diseases by 2015
7. Ensure environmental sustainability	Integrate sustainable development into country policies and programs and reverse the loss of environmental resources Halve the proportion of people without sustainable access to safe drinking water and basic sanitation by 2015 Achieve a significant improvement in the lives of at least 100 million slum dwellers by 2020
8. Develop a global partnership for development	Increase aid levels and improve aid quality Develop an open, rule-based, predictable, nondiscriminatory trading and financial system Deal comprehensively with developing countries' debt Implement strategies for youth employment Make available the benefits of new technologies

Source: United Nations (2006).

to poverty reduction by increasing the quantity and quality of their aid and by levelling the playing field of the global market through debt reduction, trade liberalization, and improved access to finance and to knowledge. This mutual accountability framework for the achievement of global poverty reduction objectives has had far-reaching consequences. By laying stress on the achievement of specific country and global results, it has raised the stakes for development evaluation. It has also shifted the evaluation focus from individual operations to systemic actions at institutional and policy levels. Finally, it has made developing countries' involvement in the evaluation process a fresh imperative of legitimacy.

The Need for Change

The implications for evaluation methods and practices are fundamental. Four major challenges now face development evaluation managers:

- First, to measure development success in terms of global goals means that the main unit of account for evaluation has shifted from the project level to the country level with new metrics derived from the commonly agreed MDGs.
- Second, to acknowledge that poverty reduction and improved governance are the primary responsibility of developing countries assumes a readiness to make development evaluation a country-based and country-owned process instead of a donor imposition.
- Third, to recognize that rich countries have an obligation to contribute to poverty reduction by improving the enabling environment for development suggests that, along with the quantity

and quality of the aid provided by rich countries, the development friendliness of their nonaid policies should be open to evaluation scrutiny.

- Fourth and finally, to commit to the MDGs implies that the realism and fairness of the goals themselves and the soundness of their underlying theories of change constitute legitimate objects of evaluation.

The development evaluation community has yet to address these challenges in earnest. Although declarations of intent about the shift toward higher-level evaluations and joint evaluations have become more frequent, most evaluators working on development are still operating in a “business-as-usual” mode. The management of evaluation resources remains firmly embedded in aid agencies. Resources allocated to evaluation capacity development remain scarce. Given the strictures of domestic aid legislations, evaluation resources are still heavily mortgaged by the fiduciary obligations associated with the design and the implementation of thousands of individual projects. Despite solemn commitments to reform aid delivery on the supply side, accountability is still one sided and directed toward the implementation responsibilities and the policy reforms required of developing countries.

Nor have the five major methodological consequences of the changed development environment been fully grasped by evaluation managers:

- The *alignment* challenge: All development interventions should be assessed in terms of their impact on the MDGs as articulated in development programs owned by developing countries.
- The *aggregation* challenge: The results of development interventions should be measured at the country level through country-based arrangements.
- The *accountability* challenge: Given the global partnership for development embedded in the MDGs, the performance of all development actors (whether governments of rich and poor countries, their civil society and private sector participants, international institutions, etc.) should be evaluated in terms of their distinctive accountabilities and reciprocal obligations.
- The *attribution* challenge: Neither the evaluation community nor the economics profession have managed to convince a sceptical public that “aid works.” More robust evidence is needed to demonstrate a causal link between aid operations and country level economic performance.
- The *asymmetry* challenge: The imbalance between the seven MDGs that address poor countries’ performance and the eighth MDG (MDG8) that addresses the need to improve the enabling global policy environment for development should be redressed; that is, the MDGs themselves (and the processes that underlie their monitoring and their oversight) should be evaluated.

Alignment

Shared objectives are a critical component of effective development partnerships, and the MDGs provide common benchmarks of operational relevance. Alignment of evaluation standards would help make a sprawling and fragmented development system somewhat more coherent by encouraging all partners to pull in the same direction. By using a common evaluation template derived from the MDGs as interpreted by country-owned Poverty Reduction Strategy Papers, alignment would contribute to convergence of evaluation incentives.

Tapping this benefit assumes of course that evaluation is carried out by relating outcomes to the achievement of predetermined goals. However, some evaluators advocate “goal-free” evaluation methods that do not proceed from a program’s official goals and ascribe costs and benefits to outcomes that are broadly consequential for the lives of participants. They argue that merit and worth are more “important evaluation predicates” than compliance with official goals because goal-free evaluation addresses efficiency and relevance of evaluated activities.¹

The solution lies in evaluating the relevance of the goals themselves—and the efficacy and efficiency of achieving them. The risks of unintended consequences involved in setting

goals and in reaching and sustaining their development impact should also be assessed. In this way, evaluation can combine accountability for program design with accountability for program results. For example, the World Bank's Independent Evaluation Group uses a development effectiveness definition that bridges the "goal-free" and the "goal-based" approaches by giving pride of place to the evaluation of relevance (in terms that are "consequential" to the countries concerned and to the World Bank), the likelihood of reaching them, and the economy of resource use in achieving them over time. The glossary issued by the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD, 2002) makes room for this approach in its alternative definition of development effectiveness as

an aggregate measure of (or judgment about) the merit or worth of an activity, i.e. the extent to which an intervention has attained or is expected to attain, its major relevant objectives efficiently, in a sustainable fashion and with a positive institutional development impact. (p. 21)

This is why the Poverty Reduction Strategy Papers have moved to the center stage of development evaluation. In principle, they translate the MDGs into development programs owned by governments and designed in consultation with the civil society, the private sector, and the development partners of the country concerned. They have become the privileged objects of development evaluation.

Aggregation

Moving evaluation to the higher plane of country results is therefore essential to connect the results of individual operations to the MDGs. This is not a trivial challenge because the sum of individual operations and of numerous country-assistance programs by diverse donors can be more (or less) than the sum of the parts. Tough methodological issues arise. Aggregation requires an assessment of the coherence of individual development operations and an understanding of the linkages among the various stages of the results chain, as prescribed by "program theory."

The task is complicated by the proliferation of aid channels and the diversity of aid vehicles. Early results from the World Bank's country evaluation experience confirm that in about one third of the cases there is a mismatch between the outcomes of aid projects and aggregate results at the country level—the micro–macro paradox. This disconnect is explained by the following factors:

- Aid funds are partly fungible, and the domestic resources liberated through aid may have been applied to other purposes (e.g., prestige projects or military expenditures).
- Where countries are highly aid dependent, aid may harm the economy by creating volatility in public revenues, contributing to inflation and raising the real exchange rate so that export competitiveness suffers.
- Excessive aid may create a "resource curse"; competition for control of rents may aggravate social tensions, reduce the incentives to reform, hinder budget discipline, and substitute donor preferences for country priorities.
- It is the aggregate impact of domestic policies and external engagement (aid and nonaid policies) rather than individual projects that matters.

Thus, the "micro–macro disconnect" confirms the need to address the aggregation challenge. This means carrying out fully integrated evaluations of all official development assistance activities at the country level. This kind of evaluation has yet to be implemented. But

there is every reason to believe that it is feasible. The experience with joint evaluation processes and products has been thoroughly examined under the aegis of the DAC. Successful experiments in joint evaluations of country assistance strategies (involving two partners) have taken place. The time is ripe for giving comprehensive, country-based evaluations the priority that they deserve.

In 2003, the chairman of DAC outlined a fourfold hierarchy of evaluations of aid effectiveness (impact of all aid on one country, effectiveness of the development cooperation system, evaluation of an individual donor contribution to the total system, and development effectiveness of an individual donor agency). Accordingly, proposals for piloting evaluations focusing on the uppermost levels of this hierarchy have been included in the work program of the DAC Network on Development Evaluation.

Finally, there is an emerging consensus regarding the approach to country-assistance evaluations. First, it acknowledges that high-quality country programs are more than a collection of disparate projects. Instead, the complementarity of projects and other aid instruments is explicitly taken into account, and the full package of financial and knowledge services provided by the donor is considered—not just projects.

Furthermore, the country evaluation seeks to identify the difference between actual outcomes and potential outcomes (without donor intervention) in seeking to determine the true value of aid at the country level. In this context, the design of relevant “counterfactuals” has proven problematic because the methodology of scenario building is not yet mature. Accordingly, a mix of program evaluation methods is adopted, including those that have long been in use in the assessment of social programs in industrial countries. This means assessing performance in relation to past results, original objectives, and outcomes of similar programs in similarly situated countries.

In evaluating the expected development impact of an assistance program, the evaluator gauges the extent to which major strategic objectives are relevant and likely to be achieved without material shortcomings. The efficacy of program implementation is measured by the extent to which program objectives are expected to be met in ways that are consistent with principles of effective aid. Efficiency ratings concern the transaction costs incurred by the donors and the country in connection with the implementation of the country assistance program. Finally, sustainability is judged by the resilience of country-assistance achievements over time, whereas institutional development impact refers to the capacity-building benefits of the country-assistance strategy.

More often than not, a triangulation of evaluation methods focused on three major dimensions is an appropriate approach:

- Evaluation of the results of individual operations, country dialogues, coordination with partners, and analytical or advisory services assessed in terms of their relevance, efficacy, efficiency, resilience to risk, and institutional impact.
- A development impact assessment, involving a “top-down” analysis of the principal program objectives and their achievements, especially as they relate to the MDGs.
- An assessment of the contribution of various actors to the outcomes according to their distinctive accountabilities and reciprocal obligations.

Although this approach reflects international financial institution experience, other development agencies use somewhat different approaches. For example, the European Union considers the impact of aid and nonaid policy vectors in assessing the relevance, quality, and size of its country program and the resulting influence on the recipient country and its partners; the Swiss Agency for Development and Cooperation emphasizes participatory techniques and

country involvement in the evaluation process; and so on. Developing countries have yet to determine their own approach to country-based evaluations, a task recently taken up by the International Development Evaluation Association.

Accountability

In principle, the aggregate performance assessment of development partners (combined with the influence of exogenous factors) is identical to the overall impact of a development intervention. However, this assumes that the relevance criterion used in judging an individual operation takes account of the degree to which coherence among development interventions is ensured. Hence, accountability assessments are inextricably linked with the assessment of results at the higher plane (aggregation). Only if they are carried out at the proper level do measurements of the performance of partners (in terms of their distinctive roles and obligations) help to combine partner accountability assessments with results-based evaluations.

In this context, the development community has begun to emphasize results-based management that stresses accountability of individual development actors both for their compliance with agreed policies, common standards, and contractual obligations and for their contribution to the delivery of measurable contributions to poverty reduction. The credibility of such assessments depends on the quality and credibility of evaluation arrangements. They require adequate resources for evaluations, relevant and accurate data collection, valid methods, and the right skills. In practice, the focus on results has not always led to greater agency accountability because results chains have not been structured to isolate the contribution of distinct actors to the realization of outcomes and impacts.

But even where results chains have been supplemented by transparent evaluations of the partnerships that characterize most development interventions, substantive threats to the independence of the evaluation process may have jeopardized its accountability dimension. The golden rule of evaluation is that it should be protected from capture by program managers and donors. Any perception that individuals or organizations affected by an evaluation can exercise leverage over the evaluation process is apt to destroy confidence in the results. Furthermore, the views of civil society partners and intended beneficiaries should be secured.

The frequent option of resorting to consultants does not guarantee independence. From a governance perspective, evaluation independence is best understood by analogy to the auditing profession. The supreme auditor reports to the legislature, and the external auditor reports to the governing board of the organization. By contrast, accountants report to the executive (or to operational managers in a corporate set up).

Supreme and external auditors are also tasked with an oversight role over internal audit and accounting functions. Finally, independent evaluation is charged with verification of self-evaluation results and attestation to a higher authority than program management that self-evaluation standards and processes are adequate. Thus, independent evaluation and self-evaluation are complementary and synergistic.

Four major criteria of independence emerge from the standards of the European Federation of Accountants, the Institute of Internal Auditors, and the Code of Ethics and Auditing Standards of the International Association of Supreme Auditors: (a) organizational independence, (b) behavioral independence, (c) protection from external influence, and (d) "conflict-of-interest" avoidance. These principles are well anchored within the auditing profession. But they have yet to be widely applied for evaluation in the governments of rich and poor countries alike. In the aid business, evaluation units directly reporting to governing boards or legislatures are found only among multilateral banks. In four bilateral evaluation agencies and the United Nations Development Program, the evaluation unit reports to the head of the

agency. In nine bilateral agencies and several United Nations specialized agencies, it has a subordinate position within a policy, audit, budget, or management division. In most development assistance organizations and in national evaluation organizations, the role assigned to local officials and to individual citizens has been marginal.

Attribution

Does aid work? The available policy research evidence is mixed. Some aid recipients have experienced growth rates that are unprecedented in world history. Although the United Kingdom took more than 60 years to double output per person (1780 to 1838), Turkey did it in 20 years (1957 to 1977), Brazil in 18 years (1961 to 1979), and China and Korea in 10 years (1977 to 1987). Between 1966 and 1990, Thailand tripled its real per capita income and India doubled its per capita income.

As for the cottage industry of statistical correlation studies, it remains inconclusive. A careful review of the evidence suggests that the effect of aid volumes on growth is small and statistically insignificant in the aggregate. However, the econometric studies that underlie this conclusion do not distinguish among aid channels, instruments, and modalities, nor do they take account of the social and institutional environment within which aid activities are embedded.

Independent evaluations suggest that more than two thirds of projects funded by the World Bank have efficiently met their relevant objectives and that the trend is up. This performance record within a public domain that is inherently risky and that offers exceptionally high rewards to global peace and prosperity does not justify the vitriolic attacks of reactionary aid critics. Of course, in sub-Saharan Africa, overall poverty rates have been rising instead of declining, and this is a region that has received a great deal of aid. Although its share of the developing world's population is about 10%, based on OECD data, sub-Saharan Africa received one third of all aid in 2004—\$26 billion out of a total of \$78 billion. But it is also a region that has endured a heavy colonial legacy and a great deal of conflict.

If, in John Lennon's words, we "imagine, there is no country," the development narrative has not been altogether bleak. As a share of the total population, poverty dropped in the aggregate between 1981 and 2001, from 67% to 53% for the \$2-a-day benchmark. Indeed, during the 1980 to 2000 period, annualized per capita growth rates were 3.6% for developing countries compared to 2.0% for rich countries.

Average social indicators in developing countries have recorded major gains: Life expectancy rose from 55 years in 1970 to 64 years in 2000, infant mortality rates dropped from 107 per thousand in 1970 to 58 in 2000, literacy rose from 53% in 1970 to 74% in 1998, and the number of people suffering from chronic malnutrition declined from 35% to 17% of the population. Aid did not do it all, but it certainly helped.

For every aid failure (and there are many, especially when aid is geared to geopolitical aims), one can point to a spectacular aid success that has delivered widespread benefits. River blindness has been banished in many countries as a direct result of aid. Clinics funded by aid have saved thousands of lives throughout the developing world. The green revolution has helped to avert large-scale famines, and it would not have been possible without the investments in new varieties, irrigation, fertilizer plants, and agricultural supporting services funded by aid. Nevertheless, the attribution issue is a critical one in the current development environment.

This is why a major effort is underway to channel scarce development evaluation resources toward experimental evaluation methods. Following a high-profile campaign of the U.S. government that enthroned randomized control trials (RCTs) as the gold standard for evaluation, an initiative promoted by the Center for Global Development, a Washington-based think tank, would spread these experimental (and quasiexperimental) evaluation techniques toward the developing world.

The methodological debate triggered by both initiatives harkens back to the controversy that rocked the evaluation world two decades ago. The controversy was harmoniously resolved when the mainstream of the evaluation community carefully delineated the conditions under which experimental and quasiexperimental methods are appropriate: They are useful only for discrete and relatively simple interventions geared to precise and measurable objectives and characterized by well-defined “treatments” that remain constant throughout program implementation.

RCTs can yield misleading results when the behavior of participants in the treatment group is affected by the experiment (Hawthorne effect). Comparison groups may also change their behavior as a result of the intervention given to the treatment group or as a result of other interventions. Nor are RCTs (given their focus on goal-based achievements) the most appropriate to deal with issues of efficiency, equity, and quality impacts. These are better captured by qualitative, participatory impact evaluations because, by highlighting differential impacts among groups, RCTs often generate public opposition even though budget constraints or other factors (e.g., political patronage) may explain the decision to limit the evaluated intervention to a specified group. Furthermore, most experimental and quasiexperimental approaches do not identify the contribution of various participants to the measured outcomes, thus blurring accountability.

Equally, the usefulness of the RCT method in a developing country context is restricted by its high cost and complexity and its heavy requirements for specialized skills. Finally, although RCTs can help assess standard, well-defined public goods delivery approaches, the method is inappropriate for broad, multicomponent country programs or policy interventions. The bottom line is that experimental methods can be used only for a very limited range of interventions that have been specifically designed to make RCTs feasible. This is why a triangulation approach that uses a combination of participatory methods, theory-based techniques, beneficiary surveys, and development case studies is better “fit for purpose.”

Asymmetry

Although the MDGs are a major achievement, it is time to evaluate and adjust them to take account of the hard-won lessons of experience. It is now widely acknowledged that most developing countries will not achieve the agreed goals. At current growth rates, East Asia alone is likely to achieve the MDGs. If present trends continue, there will be 140 million underweight children in 2020, and 75 million children will be out of school in 2015. To reach the income–poverty goal, growth in developing countries would have to be more than twice the levels achieved in the 1990s for the following 15 years.

The resources mobilized to implement the MDGs have been woefully inadequate. Probing questions about the asymmetries of the Monterrey compact have been raised. Specifically, the MDG framework demands far more of developing countries than of developed countries. Most of the agreed indicators (35 out of 48) point south. Vast resources have been mobilized to monitor progress in developing country policies and programs. No similar effort has been put in place to monitor the improvement of policies adopted by rich countries.

Serious problems of aid quality on the supply side are rarely taken into account when outcomes are evaluated. Yet tied aid increases the cost of goods and services procured under aid by one fifth. It accounts for more than half of aid flows. Furthermore, technical assistance of dubious utility accounts for almost 30% of aid. The proliferation of aid channels and its fragmentation across hundreds of small projects impose high transaction costs. Aid coordination remains a problem. Harmonization of aid practices is proceeding at a snail’s pace.

These supply-side issues have been sorely neglected in development evaluation. This asymmetrical focus may help explain why there is so much despondency about aid and why the aid industry has been so slow to reform. As importantly, development evaluation resources have been disproportionately allocated to aid. It is only one of the transmission belts of globalization (aid) and not the most important one.

There is little doubt that the governance and policy weaknesses of poor countries need to be remedied. But rich countries also need to assume their responsibilities. They exercise control over the institutions that oversee the global economy. It is their rules and their standards that regulate the flows of capital, people, and ideas. It is their production and consumption patterns that pose the greatest threat to the global environment.

Only MDG8 addresses the responsibilities of OECD countries, and it is couched in the broadest possible terms and is not buttressed by time-bound undertakings and performance indicators. Poor countries must comply with the obligation to prepare poverty reduction papers for review by the international financial institutions (as a prerequisite for debt reduction and increased aid). By contrast, rich countries are not tasked to report on their own obligations to implement the MDGs—not even MDG8.

Development Evaluation Beyond Aid

To rethink development evaluation so that it addresses global policy issues means recognizing that development evaluation has been asymmetrical, that it has focused a disproportionate share of its resources on aid, that it has neglected the other major transmission belts of globalization, and that it must reconsider its priorities and its modalities.

Leveling the playing field of evaluation would help resolve a fundamental methodological flaw. Just as project-level results cannot be explained without reference to the quality of country policies, country-level evaluations are incomplete without reference to the international enabling environment. During the 1980s and 1990s, the development evaluation community concluded that national policies in poor countries have a crucial impact on aid outcomes.

Accordingly, aid managers acted on this finding by promoting policy reform. At the turn of the millennium, the same logic should be applied at the higher plane of global policy. Yet donors and evaluators are still lavishing disproportionate attention to one side of the development partnership, the performance of poor countries. This unbalanced approach to accountability has twisted the priorities of development evaluations and undermined their quality.

Beyond aid, the quality and focus of private investment require far more independent evaluation than they have been receiving. Aid matters a great deal, particularly for sub-Saharan Africa, but it is only one of the policies of the rich countries that should be tracked. Private capital flows toward developing countries are now almost 3 times far larger than aid flows. Evaluation of trade regimes has also been neglected, even though, on a per capita basis, developing countries' exports generate more than 30 times as much revenue as aid, 12 times in the case of the least-developed countries. The protection of intellectual property rights is another neglected evaluation issue.²

The relevance of migration to the economic prospects of developing countries is rising, and again this is an issue that has not received adequate attention by evaluators.³ Finally, the policies of rich countries (emulated by poor countries) are inducing unprecedented pressures on the physical environment. Climate change threatens the most severe and widespread impacts. Africa's food security is likely to be set back. Equally, fisheries resources are being depleted because of massive subsidies by rich countries (\$15 billion to \$20 billion a year).⁴

Finally, security policies need objective evaluation. Military expenditures are 20 times larger than aid outlays. Yet they have not delivered security, and conflict remains a major obstacle to development. Of the 40 poorest countries, 24 are in or emerging from war. There are 10.4 million refugees worldwide, and 20 million to 25 million displaced people. The end of the cold war brought some proxy wars to an end, but new wars started and the carnage continues. Since the end of the cold war, the world has seen 58 armed conflicts in 46 locations, most of them in developing countries. About 300 million small arms are in private hands (more than half the world total of 550 million). Some 300,000 children have been compelled to bear arms and fight in 36 countries. Land mines perpetuate the violence of past conflicts, causing 15,000 to 20,000 victims a year.

Independent evaluation should also be promoted in the private and voluntary sectors because they have become increasingly active in the developing world. Already, rating agencies have been developing indicators to help channel capital toward environmentally and socially responsible uses. Private companies have begun to adopt “triple bottom line” concepts and to make independent arrangements for independent certification to assess the social and environmental sustainability of their own investment performance. Greater transparency in reporting is being adopted voluntarily and/or legislated. Pension funds are increasingly moving their investments out of companies with poor social and environmental records. Nongovernmental agencies have also started to enhance their own accountability systems.

Toward New Development Evaluation Priorities

The challenges reviewed above (alignment, aggregation, accountability, attribution, and asymmetry) add up to a demanding evaluation reform agenda. First and foremost, the center of gravity of evaluation should move from north to south and from west to east. This will require a major evaluation capacity-building effort and a deliberate shift of evaluation resources toward developing countries and transition economies. Poor countries’ evaluators should increasingly be put in the driver’s seat of development evaluation. Thus, the fragmentation of evaluation products would be overcome and “evaluation bombardment” would be avoided. More resources also need to be devoted to nurturing of convergence in evaluation methods through networks of evaluators, joint or parallel evaluations, and country-based evaluations connected to enhanced processes of public expenditures management.

Second, regarding alignment and aggregation, the inadequate organizational incentives and/or low priority given to monitoring and evaluation in development operations should at long last be overcome. At the country level, substantial progress has been made in the generation of household surveys and poverty data, but evaluation has yet to address poverty-reduction strategy processes in a systematic fashion. This should be done based on explicit theories of change with emphasis on public expenditure evaluations, logical frameworks, tracking surveys, participatory methods, and results chains that help to align country program outcomes with the MDGs. For evaluation of aid delivery generally, the solution lies in country ownership, harmonization, coordination, and forging principled partnerships that specify the obligations of all parties.

Third, regarding the accountability dimension, results should be connected to individual actors according to their distinct accountabilities and reciprocal obligations. This hinges in large part on the independence of evaluation arrangements and the participatory content of evaluation methods. To achieve credibility and effectiveness, evaluation processes should combine independent evaluation, self-evaluation, and full involvement of citizens—the intended beneficiaries of development programs. At its best, self-evaluation should be participatory and integrated in business processes, public administration, and partnerships.

Independent evaluations should be handled at arm's length from program managers and operational managers in funding agencies. This would help to ensure that evaluation contributes to accountability, learning, and adaptability. Self-monitoring and self-evaluation should be built into all transactions at the project, country, and global levels, whereas independent evaluation should be tasked with attesting to its rigor and quality. Thus, the setup of professional and independent evaluation units that report to country legislatures or governing bodies or that operate as independent verification mechanisms in the civil society is a top priority. Currently, this prerequisite of evaluation credibility is missing in most governments, companies, voluntary organizations, and development agencies.

Fourth, regarding attribution, more recourse to impact evaluations will be necessary, and, where appropriate, experimental and quasiexperimental methods should be used. But here as elsewhere, the nature of evaluation questions should drive the choice of methods rather than the other way round. To combine rigor with cost-effectiveness, triangulation of methods will often be the answer.

Finally, meeting the asymmetry challenge will mean filling a major evaluation gap at the regional and global levels where monitoring and evaluation are largely absent. Collaborative programs designed to deliver global public goods are currently not subjected to independent appraisal and, as a result, often lack clear objectives and verifiable performance indicators. In addition, the impact of developed-country policies on poor countries is not systematically assessed, even though aid, debt, foreign investment, pollution, migration patterns, and intellectual property regimes are shaped by the decisions of developed-country governments. For globalization to have a human face, the international rules and standards governing the global marketplace, the poverty impact of trade, migration, intellectual property, and other rich countries' policies, and the diverse collaborative public goods delivery programs currently underway will have to be evaluated as systematically as aid projects have been.

Conclusions

Tackling this agenda will require a redeployment of resources away from the evaluation of individual interventions toward (a) evaluation at a higher plane, (b) evaluation capacity building, (c) citizens' participation in evaluation, (d) evaluations of the mutual accountability between aid donors and recipients, and (e) evaluation-research partnerships designed to capture the poverty impact of policies and standards shaped by rich countries and the policy networks that they control.

In sum, a new paradigm has transformed the evaluation environment and raised the bar for the fledgling development evaluation profession. In the new century, development evaluation will have to be transformed and liberated from its current strictures to reach beyond aid, beyond projects, and beyond top-down approaches. It should be more comprehensive, more participatory, and better adapted to the felt needs of society.

In the new development environment, the five As (alignment, aggregation, accountability, attribution, and asymmetry) will have to be tackled. This will require principled partnerships that level the evaluation playing field, treat developing countries as full owners of evaluation methods and processes, and ensure fair and rigorous assessment of regional and global policies and programs.

To move development evaluation to this higher plane, a new evaluation architecture will have to be erected so as to give a privileged role to country-based evaluations with developing countries in the driver's seat. Finally, a pluralistic approach to methods will have to be adopted. Just as form follows function, the nature of the problems being evaluated should govern the choice of methods rather than the other way round.

In their handbook (*Evaluation for the 21st Century*) produced a decade ago, Eleanor Chelimsky and William R. Shadish (1997) wrote, "Evaluation is becoming international in the sense of being at the same time more indigenous, more global and more trans-national" (p. xi). They were prescient: Evaluation has become more global and more local all at once. They went on to note that "evaluators will be called upon to exhibit considerable courage in the normal pursuit of their work" (p. xiii). They were right: Without courage, all evaluation instruments are blunt. Although evaluation should identify and celebrate principled behavior and effective interventions, it should also uncover and disseminate knowledge about failed programs, incompetent practices, and faulty policies. This is not a popular thing to do. But it needs to be done.

Notes

1. The debate between proponents of "goal-based" and "goal-free" evaluation illustrates the crucial difference between traditional auditing and evaluation: The former emphasizes compliance with agreed objectives or standards, whereas the latter goes beyond this to objectively assess the objectives and standards themselves, including the cost of achieving them.
2. Trade-related intellectual property regimes have restricted access to essential drugs and other knowledge-intensive products and services. There is increasing pressure on developing countries to increase intellectual property protection based on the current practices of developed countries. Yet standards that may be suitable for developed countries produce more costs than benefits in poor countries that are large net importers of technology.
3. With goods, capital, and ideas more mobile, migration makes eminent economic sense. Remittances are 5 times the amounts of aid to Latin America and the Caribbean. They account for one fifth of Jordan's national income. They are the largest foreign exchange earner of El Salvador, Honduras, and the Dominican Republic. Current immigration policies obstruct the entry of asylum seekers, interdict entry by unskilled migrants, and deliberately ration immigration toward well-trained professionals and skilled workers in high demand. Such discriminatory policies induce a "brain drain" from poor to rich countries. Albania has lost one third of its qualified people. Also, 30% of Mexico's PhDs and three fourths of Jamaicans with higher education live in the United States.
4. Industrialized countries dominate global environmental management through the heavy ecological footprint of their production and consumption patterns and indirectly through their influence over global regimes governing trade, investment, and the global commons.

References

- Annan, K. (2002). Problems without passports. *Foreign Policy*, No. 132.
- Chelimsky, E., & Shadish, W. R. (1997). *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage.
- Organisation for Economic Co-operation and Development. (2002). *Glossary of key terms in evaluation and results based management*. Paris: Development Assistance Committee.
- United Nations. (2006). *The millennium development goals report 2006*. New York: Author.