Evaluation

SAGE Publications (London,
Thousand Oaks and New Delhi)
[1356–3890 (200207)8:3; 322–339; 027743]
Vol 8(3): 322–339

The Logic of Mainstreaming

A Development Evaluation Perspective[1]

ROBERT PICCIOTTO
The World Bank

The rationale for evaluation mainstreaming is straightforward: evaluation creates value only when lessons are drawn and incentives encourage institutional learning. Accountability as well as organizational adaptation are facilitated by the incorporation of evaluation processes in policy formation and program administration. This article is about the 'what, why and how' of evaluation mainstreaming; it does not dwell on the 'whether'. First, mainstreaming is defined and its prominence in the policy-making discourse is explained. Second, a theory of mainstreaming is suggested and its relevance to the development process in poor countries is highlighted. Third, the process of evaluation mainstreaming within organizations is examined, using the World Bank as an example.

KEYWORDS: effective evaluation; organizational learning; policy design; program management

Mainstreaming vs the Mainstream

The term 'mainstreaming' has become so fashionable in public policy circles that the notion, once rich in promise, has become trivialized through repeated use. It is time to restore clarity to the concept. Dictionaries are of little help since they provide definitions for the noun (*the mainstream*) but not for the verb (*to mainstream*). The mainstream is the 'principal course of activity' or the 'major current of opinion'. The first part of the word (*main*) connotes dominance and constancy. The second (*stream*) has placid even bucolic undertones. Given its aquatic origin, the term evokes fluidity, harmony and inevitability.

To adopt mainstream attitudes is to 'go with the flow' and avoid exertion against the social forces of gravity. To espouse mainstream views is to minimize the risk of confrontation and ostracism. To endorse mainstream policies and programs is to benefit from social approval. The halo of these peaceful characteristics extends beyond the noun and colors the perceptions associated with the verb. This is not entirely justified: just as the tranquil surface of a large body of water can conceal a great deal of turbulence, the organizational mainstream shifts with the ebb and flow of the coalitions that sustain it.

Fed by many tributaries, the mainstream is a composite of competing views and diverse beliefs. Its undercurrents can be tumultuous. For example, the evaluation mainstream comprises many schools that vigorously contend for leadership. Similarly, the mandate of the World Bank (to promote growth and poverty reduction in developing countries) is a composite of the highly diverse views of a membership that spans all regions of the world.

The verb (*to mainstream*) is a dynamic concept. It suggests a deliberate perturbation in the natural order of things. It creates winners and losers, challenges vested interests and triggers changes in alliances. It subverts the status quo and yet it does not evoke chaotic change or painful disruption. In effect, mainstreaming connotes gradual reform rather than frantic revolution. In policy terms, it is typically achieved through incremental changes in program goals, protocols of operations and organizational cultures.

For example, the mainstreaming of environmental considerations in World Bank operations was initially driven by outside advocacy groups. They skillfully influenced public opinion, used the media, orchestrated political pressure and sought allies within the organization to promote their agenda. The strategic shift took many years to accomplish and it ultimately induced far-reaching changes in organization, core competencies, lending procedures and business partnerships.

The impact of this policy shift remains controversial, e.g. the World Bank's 1991 forest strategy brought the conservation agenda within the mainstream of World Bank operations but it produced a chilling effect on World Bank involvement in sustainable forestry management in the tropics. An independent evaluation by the Operations Evaluation Department facilitated a reconsideration of the underlying forest strategy (World Bank, 2000a). In general, full mainstreaming of environmental sustainability in Bank operations remains a challenge (World Bank, 2001a). Gender mainstreaming has proven even more so (World Bank, 2001b).

Why is Mainstreaming so Elusive?

Mainstreaming brings to the surface the turmoil of the deep. Whereas the mainstream is the old order, mainstreaming seeks a new order. Shifts in program direction, implementation modalities or resource redeployment are easier to announce than to achieve. Public support for new policy directions is not automatic. Stakeholders negatively affected by a new initiative will inevitably oppose it. By its very nature, mainstreaming (unless it is merely rhetorical) threatens the mainstream.

While the term mainstreaming evokes ease and inevitability, the actual process is riddled with difficulties. Belonging to the mainstream only requires stalwart loyalty. By contrast, to bring a mainstreaming initiative to fruition calls for innovation and adaptability combined with consistency of purpose and the ability to consolidate support for change – a rare combination. To be sure, the mainstream is often made up of subsidiary currents and countervailing forces that can be redirected or mobilized to facilitate mainstreaming. Thus, mainstreaming:

> . . . works by a series of secessions from the dominant culture. But also – and this must be understood – it works by a series of introjections into the dominant culture. No statistics can measure the effect of these secessions and introjections but clearly the effect is very considerable. (Trilling, 1947)

But tolerating dissent within an organization (e.g. by endorsing the creation of an independent evaluation group within it) is not common. It enhances institutional resilience – just as flexible engineering designs increase the durability of structures subjected to shocks. But it also presupposes that the prerequisites of a learning organization are in place. In general, mainstreaming is facilitated by organizational and social tolerance of dissent and debate; but there are limits. Just as river mechanics demonstrate that channels can be vulnerable to breach depending on the volume and turbulence of flows, organizational and social cohesion can be undermined by irreconcilable differences in social tenets, organizational values or behavioral norms. In the face of determined opposition, mainstreaming sponsors often back down or settle on symbolic actions designed to placate opinion rather than to induce major change.

Thus, a policy maker may announce that a program has been adopted without making the hard decisions necessary to implement it. Such tactics can be discouraged through independent evaluation, but the introduction of independent evaluation itself may be thwarted through capture or manipulation. It follows that a careful assessment of readiness to change is critical to the judicious design of a mainstreaming intervention (Box 1).

Rewards and Risks of Mainstreaming

In today's volatile economic environment, frequent mainstreaming of new concepts and lessons by the society is a necessity. To retain their effectiveness, policies, programs and organizations need to be adjusted periodically. For example, as the global economic environment changes, public policies and programs must reposition their boundaries, merge with other public initiatives or restructure their goals and instruments. Facilitating this process is the major task of World Bank assistance to developing countries.

Box 1. Assessing Ownership of Policy Change

An independent evaluation of policy adjustment lending concluded that actual implementation of measures agreed by government authorities, i.e. their mainstreaming in the body politic, hinged substantially on broad-based ownership of the operation. The degree of such ownership was measured using the following indicators:

- degree of participation in the design of the operation;
- intellectual conviction reflected in actions taken upstream of loan approval;
- degree of consensus of the national leadership; and
- use of participatory processes to secure public support for the operation.

Source: World Bank, 1993

This process can be demanding since the very changes required to adapt to new external conditions may break valued traditions, disrupt alliances, weaken bonds of stakeholder loyalty and eventually undermine the very stability on which program effectiveness depends. To be sure, lack of change may be riskier than change. For example, given globalization, the economies of countries that do not hook up to the mighty engine of the world economy will languish. On the other hand, those that do are frequently required to reform their policies and institutions. This may cause economic hardship and social tensions as well as reform backsliding – as the tortuous progress of policy reform in developing countries demonstrates.

While the benefits of mainstreaming can be great, there are asymmetries in risks and rewards. The principals of a mainstreaming initiative (those who have the authority and the responsibility) may rig the 'rules of the game' to benefit particular groups. A decision to mainstream may be a stepping stone for further advancement and increased public approval. For the agents of mainstreaming, 'going with the flow' and conforming to the desires of the hierarchy can produce rewards for loyal service. However if the initiative fails, careers may be harmed, reputations damaged and livelihoods threatened.

Thus, agents directly associated with the genesis and implementation of a mainstreaming initiative may benefit disproportionately from its successful implementation while failure may lead to spectacular reversals of fortune for those visibly associated with the initiative. This explains why a 'wait and see' attitude is often prevalent, especially at the initial stage of a mainstreaming initiative. In bureaucratic settings, being right too early is simply another way of being wrong.

Management of the Risks of Mainstreaming

Just as a river diversion involves major civil works, to mainstream a new operational emphasis requires substantial investments. The costs are immediate and direct while the benefits are prospective and indirect. The mainstreaming process requires incremental resources, the exercise of compulsion and persuasion, changes in incentives, adoption of new procedures, adaptation of training protocols, etc.

As noted by Machiavelli, those negatively affected by a 'change in the order of things' will oppose it vigorously while those who may benefit from it will be restrained in their support. Therefore, to mainstream a new policy direction requires that individuals entrusted with the execution of a development intervention exercise imagination, judgment and courage but also shrewdness and political nimbleness. They must take charge, take risks and remain ahead of the curve. Mainstreaming is a key function of leadership.

Thus, mainstreaming managers must strike an appropriate balance between conformity and diversity as well as between stability and change. This is not a frequent achievement. According to Victor Hugo, the rare art of governing consists in putting just enough of the future into the present. Successful organizations are those that continually shape their policies and structures to fit their evolving markets (Milgrom and Roberts, 1992). Similarly, national economic policies must adapt constantly to shifts in the volatile and integrated global economy.

To perform efficiently, internal transaction costs must be kept low through trust and predictability while program adaptation is needed in order to reflect an operating environment in perpetual transformation. For example, World Bank experience in Brazil has shown that economic growth and demographic patterns are responsible for an epidemiological transition which calls for a fundamental reorientation of public health strategies (World Bank, 1999).

Towards a Theory of Mainstreaming

Mainstreaming of significant initiatives requires fundamental changes in patterns of opinion, influence and activity. Hence, the need to connect mainstreaming to theories of behavioral change in society. I will consider here three distinct paradigmatic approaches (Granovetter and Swedberg, 1992). The *first*, championed by sociologists, perceives human beings as highly responsive to authority and pliable to the opinions of others so that mainstreaming mostly requires decisions by leaders within the context of social customs, values and norms. The *second*, held by economists, views social systems as atomized and made up of individuals motivated by self-interest. The *third*, transcending the other two, puts the focus on formal and informal relations. Articulated by institutional economists (and economic sociologists) it views human action as 'embedded' in a web of information networks and social links.

These rival paradigms are complementary. Each offers useful insights both for mainstreaming and for evaluation. The *first* paradigm is *hierarchical*. It holds that individuals operate under the sway of predetermined, socially validated protocols. Once located in a particular time, class and group, the behavior of an individual is fixed autocratically and/or through custom. This conception is consistent with Hobbesian theory (greatly favored by dictators and CEOs) according to which a chaotic and violent 'state of nature' must be domesticated by a sovereign authority within the constraints of custom, tradition and law. Under this model, mainstreaming depends largely on the decisions of the leader. It favors 'client-based' self-evaluation.

The *second* paradigm of societal change is *individualistic*. It gives primacy to self-preservation, self-interest and the drive for economic advantage. It postulates self-centeredness and rational choice based on personal assessments of the costs and benefits of alternative courses of action. Given such assumptions, mainstreaming requires only that adequate information is provided and judicious incentives are set to ensure that individuals act in the desired manner. Within the firm, personnel evaluations, score cards and other results-based management instruments are associated with this view of behavioral change. Consistent with this model, product testing is the relevant evaluation discipline for consumers while in the political arena, independent policy evaluation serves to guide voter choice.

To the extent that the interests of individuals are consistent with the general welfare – as postulated by Adam Smith's invisible hand – it is sufficient to demonstrate that a new policy is in the personal interest of a majority of individuals to achieve its mainstreaming. In reality, of course, the invisible hand only operates where perfect information and perfect competition prevail. Catering to one's private interests does not automatically enhance the general welfare. Some institutional contexts create harmony between individual and societal ends. Others do not since collective action is frequently distorted by self-interested behavior and 'free rider' problems.

Hence, the emergence of a *third* paradigm which emphasizes hybrid mixtures of co-operation and competition. This view of how societies function is *relational*. It stresses the role of formal social structures but also of personal links and informal networks. It views individual and collective action as embedded in social relations. It recognizes that major social achievements and disruptions result not from individuals acting alone but through networks of influence, coalitions and a wide range of interlocking formal and informal organizations. This paradigm perceives negotiations among groups (rather than hierarchy or individual choice) as the key to social change. It views mainstreaming as an ongoing process of mediation among conflicting interests. The bargaining involved is helped along by empowerment evaluation and evaluation as argumentation and negotiation.

These three paradigms of society form a continuum. They are complementary and each has something to contribute to mainstreaming and to evaluation. Rather than opting for a single paradigm, successful mainstreaming and dissemination of evaluation results involves a mix of instruments that combines the insights of the three approaches. The *hierarchical* paradigm, served by client-based self-evaluation, promotes mainstreaming strategies based on corporate values and mission statements, loyalty symbols and good old-fashioned compulsion. The *individualistic* paradigm is consistent with evaluation testing techniques used for assessing personnel performance and tracking staff attitudes. It contributes to corporate management through performance auditing, results-based contracting, budget incentives, etc. Finally, the *relational* paradigm emphasizes the use of networks of knowledge and influence and gives pride of place to empowerment evaluation and self-evaluation. It focuses on specific social contexts, assesses social mechanisms and addresses unique situational constraints and opportunities.

Other hybrid evaluation schools have emerged. Thus, *participatory evaluation* lies at the intersection of hierarchy and autonomous interest groups while *comprehensive auditing* is the child of hierarchy and individual choice and *realistic evaluation* takes account of individual motivation mechanisms as well as of social relationships. Figure 1 summarizes the relationships between these various contemporary evaluation movements and social science paradigms.

Since evaluation is a meta-discipline, evaluation practitioners are understandably impatient with the paradigmatic conflicts that often rage among the social sciences. Conversely, decision makers have limited tolerance for the family fights frequently witnessed among evaluators of different persuasions. The truth is that each evaluation school has something distinctive to contribute to the progress of the evaluation profession. Accordingly, mainstreaming *of* evaluation needs mainstreaming *within* evaluation.

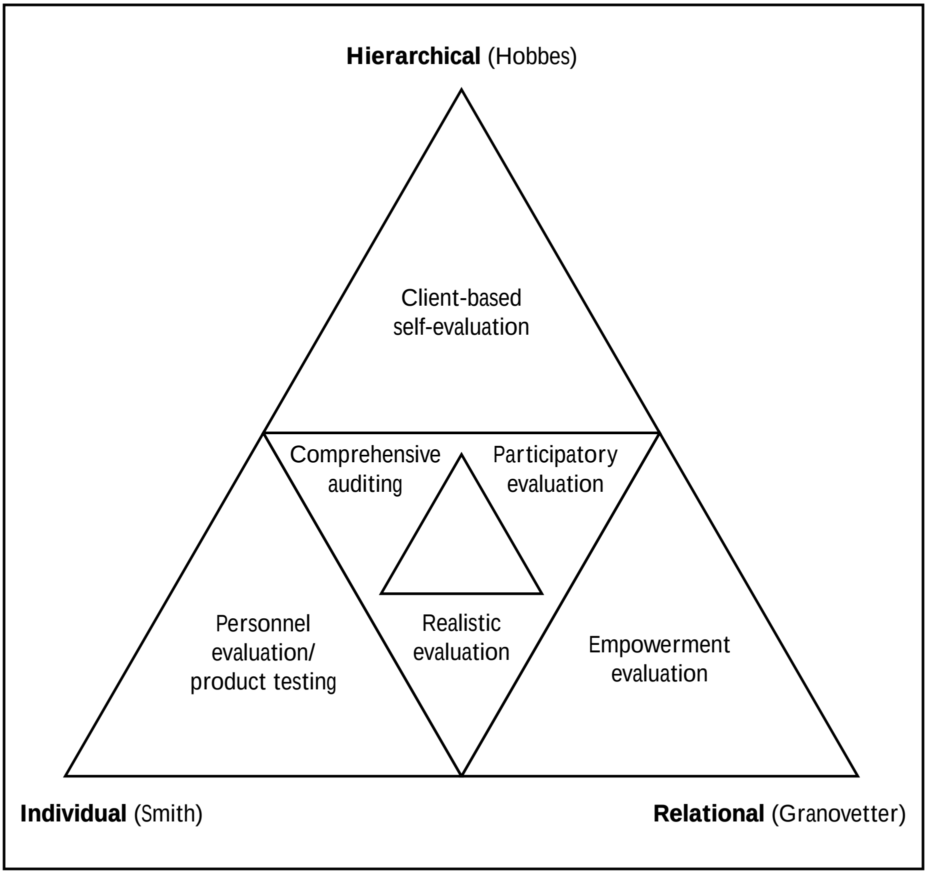

Figure 1. Social Science Paradigms and Evaluation Schools

Combining the Paradigms

The 'tipping point' theory (Gladwell, 2000) illustrates how the hierarchical, individual and relational views of society can be combined to explain mainstreaming. It deals with:

> . . . the emergence of fashion trends, the ebb and flow of crime waves or, for that matter, the transformation of unknown books into best sellers, or the rise of teen age smoking, or any number of the other mysterious changes that mark everyday life. (Gladwell, 2000)

The 'tipping point' is reached when mainstream opinion shifts and new ideas become the general consensus. Three types of social actors intervene to transform the world through word of mouth: connectors, mavens and salesmen. *Connectors* trade on the 'strength of weak ties' elaborated by Mark Granovetter (1973). They inject new ideas within an organization or the society at large through their easy access to large numbers of influential friends and acquaintances. *Mavens* know things that the rest of us don't. They specialize in accumulating and disseminating accurate and timely knowledge. Finally, *salesmen* have the intuition, the charm, the energy and the motivation to cajole, persuade and transform knowledge into action. They generate emotional contagion.

The Gladwell thesis explains the paradox highlighted by Carol Weiss among others (see Alkin, 1990) according to which good evaluations may fail to have an impact while those that are mediocre sometimes do. Beyond their intrinsic validity, the key to effective mainstreaming of evaluation lessons has to do with the kind of channels which evaluators use for the transmission of their findings. Typically, evaluators are mavens. They are most effective when they strike alliances with individuals, groups or networks with complementary connectivity and influence assets. Using such leverage, evaluators can spread wholesome 'epidemics of knowledge' and help achieve worthwhile mainstreaming objectives.

Mainstreaming and Development

In the development assistance business, mainstreaming means the widespread adoption of a new policy, a new approach to the delivery of public services or a new method of program management, taking full account of the country context. The overwhelming preoccupation with policy adjustment and institutional change is relatively new. Fifty years ago, when the development assistance enterprise was launched, it was conceived mostly as a process of physical capital accumulation. Accordingly, the task of development finance was mostly a matter of transferring resources to developing countries through the vehicle of investment projects. Gradually, the concept of capital was expanded to incorporate knowledge, protection of natural resources, institutional capacity and social relations.

For example, in 1996, the World Bank decided to support a primary education improvement project sponsored by Chile's democratically elected government. It was designed to enhance the quality of instruction in government schools, especially in rural areas, in order to help overcome the social imbalances associated with a voucher program implemented under the previous administration (World Bank, 2001c). The project employed a mainstreaming strategy involving support for an expanded pre-school scheme, the use of enriched instructional materials, participatory-education training and computer-literacy programs. The project led to a reduction in the achievement gap between low-income schools and the national average (from 30% to 10%).

A comprehensive conception of development now reigns within the development community. Its strategies aim at nothing less than the wholesale transformation of society. Today, through mainstreaming of policy change and institutional development, World Bank operations are leveraged to influence the overall development patterns of the economy. For example, following the demise of communism in Eastern Europe, a US$120 million loan was granted to Romania to help reform its transport institutions, the upgrading of Romania's national highways, the commercialization of road maintenance and the improvement of road financing systems (World Bank, 2001d).

The gradual expansion of the development agenda is itself an example of mainstreaming. The comprehensive development framework (see Figure 2) is being piloted in twelve countries with World Bank support. This framework is a set of principles proposed by the President of the World Bank to improve the quality of aid to developing countries. It relies on four key principles:

- a holistic, long-term perspective;
- partnership between government, the private sector and the civil society;
- government commitment to policy reform;
- results orientation.

It is emblematic of the multi-faceted, multi-disciplinary and holistic conception of development sketched above. It combines a long-term development perspective, partnership among all development actors, ownership of policy reform and institutional development and results orientation.

Viewed through this lens, mainstreaming is development. As for all other

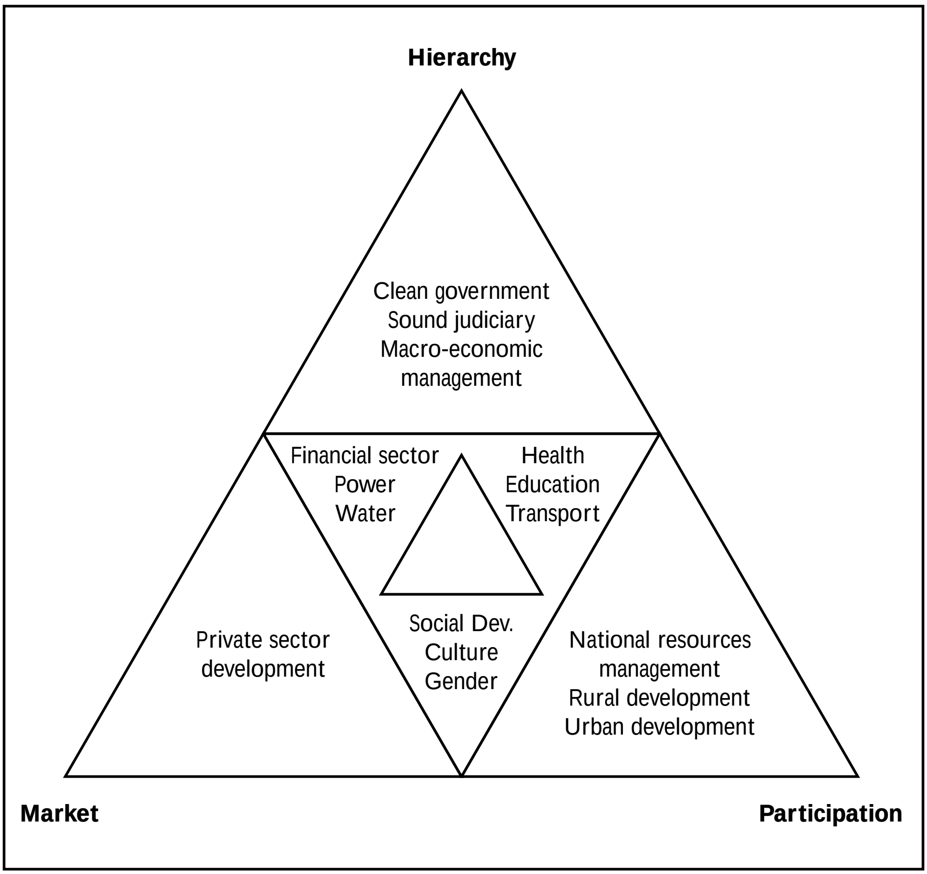

Figure 2. The Comprehensive Development Framework

mainstreaming initiatives, development operations cannot overcome path dependence in an organization, let alone an entire society, without threatening some of those who support the status quo. It may even require the questioning of deeply ingrained values, customs and protocols. Thus, the adoption of market-friendly policies may require a rethinking of the role of the state in society. Similarly, involving women in development activities may be viewed as subversive of domestic cultural values. This is why the World Bank's mission of facilitating the adaptation of poverty stricken societies to changes in demography, technology and global market conditions is so complex and controversial.

To pursue the hydraulic metaphor of mainstreaming, the training of rivers and the design of embankments to avoid flooding in populated areas are intricate engineering tasks, especially where rivers are unstable. Similarly, in poor countries with limited institutional capacity, far-reaching development rewards involve risks so that not all policy and program experiments succeed. Indeed, there are situations where the institutional and policy environment is so poor that countries cannot absorb development assistance effectively. How to inject policy change and build capacity for institutional reform in such situations is a frontier issue for development assistance.

In Kenya, the World Bank sought to improve the management of extension services starting in 1982 by introducing a disciplined program of farm visits and research-based training for extension agents (World Bank, 2000b). Funding allowed the establishment of an integrated national system, improved skills and better linkages between extension and research. The system expanded rapidly but mismatch between farmers' needs and extension advice, lack of private- and voluntary-sector involvement, excessive staffing in relation to budget resources and a paucity of new technology packages led to declines in productivity so that the program eventually became financially unsustainable. It proved impossible to reverse the decline of the extension services given the general deterioration in Kenya's civil service quality and governance.

Thus, one cannot judge the efficacy of a particular form of mainstreaming without an accurate appreciation of the context within which it is being attempted. For example, efforts to mainstream gender considerations in development assistance programs cannot be expected to succeed merely by introducing monitoring and reward systems within the donor agency since results can hinge on whether women's rights are protected by law and custom in the recipient country. Nor should one expect to mainstream compulsory universal education in rural schools, merely by influencing community leaders. For families living at the margins of subsistence there are compelling economic incentives to have children tending livestock or helping in cultivation instead of attending school. Only a sequential and progressive attack on the generic obstacles that prevent the achievement of priority social goals can achieve durable results.

The marshalling of social forces and individual energies which follows a crisis of confidence in a public venture typically involves the injection of a new policy direction through the body politic or the replacement of a faltering program by a more promising intervention. But mainstreaming of one or more policy or program priority does not take place in a vacuum. In order to succeed, the mainstreaming mechanism must take account of the specific causes of performance setbacks, societal disappointments or organizational lapses and deal with each of them. This means a withdrawal from a prior unsatisfactory arrangement followed by fresh involvement by participants in the design and the implementation of new arrangements. In other words, successful mainstreaming implies a willingness to *exit* from past entanglements and to activate the *voice* of relevant stakeholders (Hirschman, 1970). The particular exit-voice mechanism selected must reflect the particularities of the context.

In the Chile education case, the World Bank helped to re-orient education policy towards more equitable outcomes, in response to the voice option exercised by poor communities. In the Romania transport case, the voice option was strengthened through deregulation, new autonomous regulatory agencies and the privatization of road maintenance. In the Kenya agriculture extension case, the World Bank exited from supporting the monolithic training and visit system so as to give farmers a voice in the extension delivery system and make more effective use of partnerships with the private sector and the civil society.

In effect, mainstreaming in development assistance involves judicious admixtures of exit (through the application of economic concepts) and voice (via the promotion of participatory processes). *Horizontal* mainstreaming involves the replication of a development project across a geographical or functional area. This is often referred to as up-scaling a development intervention. *Vertical* mainstreaming, on the other hand, aims at the incorporation of one or more new policy objectives within an existing organization or program.

An example of *horizontal* mainstreaming is the expansion of the dairy co-operative movement in India funded by the World Bank and other donors (World Bank, 1998). It started in a single district of India but, with external support, gradually turned into a nationwide program benefiting small producers by offering them access to modern dairy processing facilities and well-run marketing networks. The combination of market discipline and social capital creation in thousands of Indian communities participating in Operation Flood is an apt illustration of the synergy between exit and voice which innovative rural development schemes can trigger.

An example of *vertical* mainstreaming is the introduction of social and environmental safeguards in World Bank operations. Here too, the exit option and the voice option have been combined. Where countries do not comply with the safeguards, loans are canceled (exit). On the other hand, the resettlement policy prescribes elaborate consultations with the populations affected by infrastructure development (voice). An Independent Inspection Panel ensures that harm to local communities as a result of alleged violation of World Bank policies is independently investigated. The record suggests that compliance at the level of the project is hard to achieve where the domestic regulatory environment is inconsistent with the requirements of the safeguard policies. Thus, vertical mainstreaming within the World Bank may require horizontal mainstreaming at the country level.

In the least developed countries, investment in human capital, voice formation, promotion of voluntary associations and creation of market institutions may be a prerequisite to mainstreaming. In such environments, institutions need to be nurtured, catalyzed and mobilized through capacity development as a necessary antecedent of mainstreaming. For example, a social forestry and forest management project in the Terai area of Nepal (World Bank, 1992) supported by the World Bank failed due to inadequate training and capacity building. A successor project aiming at forest protection in the hills delivered better results because it explicitly provided for the establishment of an institutional framework and relied on community participation instead of target-driven implementation blueprints (World Bank, 2001e).

The diversity of the development interventions sketched above suggests that different instruments are needed to achieve mainstreaming. Here too, a triangular depiction is helpful to delineate the distinctive and complementary characteristics of each instrument within the World Bank's toolkit (see Figure 3). The investment project is the workhorse of Bank operations. It reflects early hierarchical conceptions of development as blueprint interventions. The adjustment loan is used to reform policies. It originated following the debt crisis of the early eighties with the realization that macro-economic distortions hindered sustainable and equitable growth. The community-based operation represents a grassroots response to the problems of poverty reduction. Bridging the first and second of these instruments is the financial intermediary loan. Bridging the second and third is the sector adjustment or investment loan while social funds are hybrids between the specific investment project and the community-based project. All of these instruments are subjected to systematic evaluation.

Evaluation Mainstreaming

At one level, the mainstreaming of evaluation is the generalized acceptance and adoption of particular evaluation concepts, methods, practices or findings. At another, it is the incorporation of the evaluation function itself into the project, program or organization being evaluated (self-evaluation). Both concepts of evaluation mainstreaming aim at a combination of individual, organizational and societal learning. To the general obstacles which mainstreaming faces (opposition by vested interests, lack of leadership, poor incentives, lack of participatory tradition, path dependence, etc.) must be added constraints that uniquely plague the evaluation function.

Specifically, at the heart of evaluation mainstreaming lies the paradox of 'obliteration by incorporation' (Merton, 1996). It holds that social acceptance of a new idea typically results in a gradual loss of recognition of its source and meaning. A virulent variety of this syndrome plagues evaluation activities. The more visible and well documented an evaluation, the greater the chances that the program managers will readily concede its value while noting in the same breath that they had already reached the same conclusions ('incorporation through acknowledgement'). This trivializes the evaluation exercise and, in the absence of follow-up evaluations, helps to ensure that the remedial actions actually required – and especially the more exacting ones – are gradually forgotten ('obliteration').

The emulation of evaluation processes and techniques by the managers of programs being evaluated (e.g. through the set up of self-evaluation activities) is

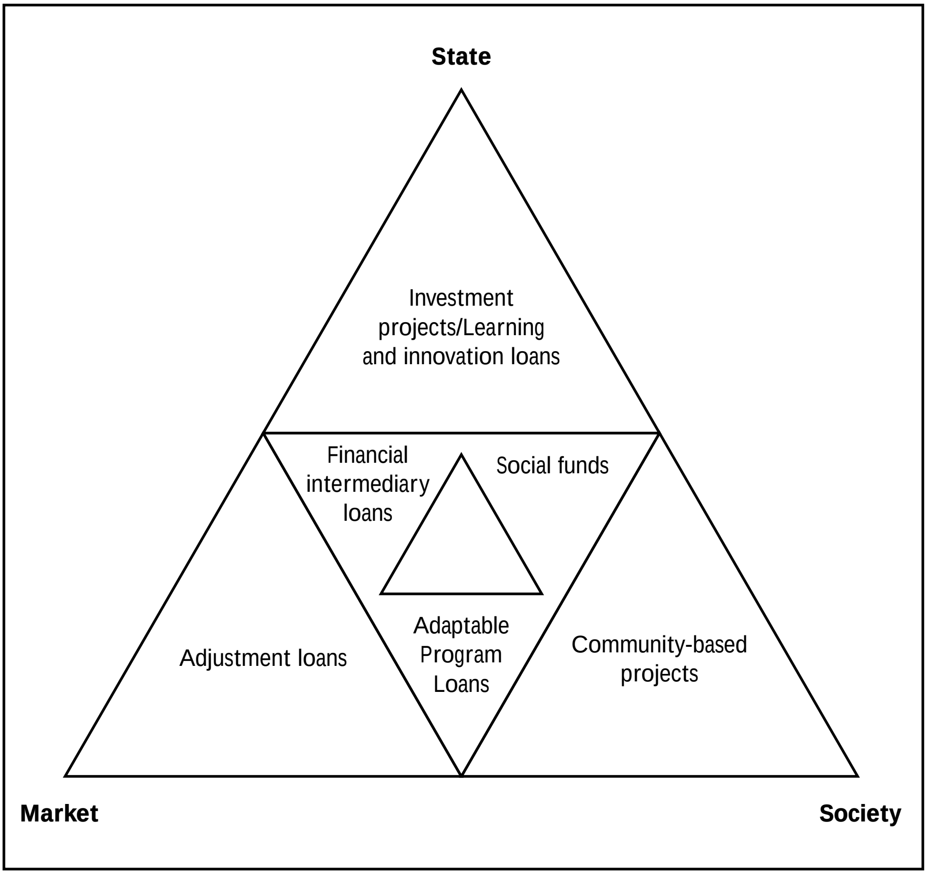

Figure 3. Lending Instruments

usually commendable and conducive to evaluation mainstreaming. But self-evaluation can be captured or used as a protective shield, e.g. where the reputation of a manager or the survival of a program is at stake. Thus, the initiation of self-evaluation studies may be used to evade independent evaluation or delay program reform. The evaluative skills assembled for a self-evaluation may be used to discredit the methods and findings of an independent evaluation so as to escape accountability. Self-evaluation may even displace independent evaluation in the organizational structure under the cover of mainstreaming, e.g. by pointing to the superior learning potential of self-evaluation and lamenting the 'redundancy' and high cost of independent verification.

The experience of the World Bank's Operations Evaluation Department (OED) suggests that workable (albeit necessarily imperfect) solutions to such problems exist. Indeed, the history of the evaluation function within the World Bank provides concrete illustrations of useful evaluation mainstreaming principles. In the fifties and early sixties, prior to the McNamara Presidency, there was no formal evaluation process except for internal auditing and end-use supervision of project execution. A formal evaluation function was created in 1970. Robert McNamara lodged it in the Planning and Budgeting Department reflecting the hierarchical management concepts and operational research techniques then in vogue.

Under the McNamara model, mainstreaming was relatively easy since evaluation was part and parcel of the control system designed by the President for his own use. It led to the adoption of utilization-based evaluation principles, many of which are still in place. But given the absence of an arm's length relationship between evaluation and line management, the system did not enjoy much credibility among shareholders. In 1974, the executive directors exercised their 'voice' option and decided that the function should become independent of management. The position of Director General, Operations Evaluation was created. Since then, OED has reported to the Board of Executive Directors of the World Bank. To protect the integrity of the function, the Director General, Operations Evaluation (DGO) cannot be removed from his post by the management and is not authorized to join the staff of the institution upon completion of his/her term.

This structure has helped to enhance the objectivity of the function and to guarantee its independence. A Committee on Development Effectiveness (CODE) of the Board oversees the workings of the evaluation system including the utilization of evaluation findings. It reviews all major OED studies and the management responses associated with each. OED compiles a management action record to track the implementation of management undertakings agreed under the aegis of CODE. The results are reviewed annually by CODE. The first DGO had the foresight of mandating self-evaluation processes while ensuring that the independent evaluation function would be endowed with the authority to attest to the adequacy of self-evaluation criteria and procedures.

This construction has endured so that independence has been achieved without incurring isolation. The evaluation governance framework has also facilitated congruence between corporate objectives and staff incentives. It has shifted the evaluation model toward decentralized accountability and learning. In an institution devoted to management by objectives, it is not surprising that goal-based evaluation was adopted or that evaluation processes were shaped to reflect well-tested auditing principles. In effect, evaluation shifted emphasis and migrated from the top of the triangle towards its left-hand corner (see Figure 1).

Throughout the seventies and the eighties, project evaluation rather than program evaluation constituted the bulk of OED's work. Since then, some 280 lending operations have been self-evaluated and independently evaluated annually following completion of disbursements. But evaluation procedures have evolved. Through a process akin to the 'creative destruction' observed by Joseph Schumpeter in capitalist economies, the introduction of self-evaluation has resulted in a constant need to reposition the focus of the independent evaluation program so that it keeps adding value to the work of the organization.

For example, as the quality of self-evaluation improved, the ratio of desk reviews to field reviews of individual operations by OED gradually increased.

This allowed OED to shift resources from direct evaluation of projects towards the higher plane of program and policy evaluation (country and global assistance programs, operational policies, etc.). In parallel, the World Bank's evaluation processes have had to evolve to adapt to changes in the organization.

The primary preoccupation with project financing has given way to a focus on knowledge creation and dissemination which has led to new evaluation methods and processes. The World Bank has built bridges towards other development agencies, the private sector and the civil society in order to 'leverage' its development impact and concentrate its services in the areas of its comparative advantage. This has triggered the use of partnerships in evaluation. Finally, the World Bank has broadened its development agenda to enrich its traditional focus on economic management by promoting social and environmental sustainability. This has induced changes in evaluation skills and priorities.

OED now seeks to provide a platform for evaluation partnerships spanning the entire development system. This has become necessary to allow systematic evaluation of collaborative global programs, co-ordinated resource mobilization for debt reduction, poverty reduction strategies in poor countries, reconstruction assistance to post-conflict countries, emergency lending in the wake of financial crises and comprehensive development framework pilots. This diversification has required changes in evaluation objects, benchmarks, techniques and processes.

For example, the mainstreaming of the new international development consensus within the World Bank's operational agenda has led OED to add institutional, social and environmental criteria to the traditional economic and financial benchmarks. Equally, a tighter connection between evaluation processes and the business cycle of the Board of executive directors has been instituted to enhance the relevance of the function (Ingram and Feinstein, 2000). Improved co-ordination between independent evaluation, self-evaluation and auditing functions has reduced perceptions of control overload.

Today, OED assesses the extent to which the policy advice, lending and non-lending services offered by the World Bank, translate into beneficial changes in development practices and patterns at project, country and global levels. Through evaluation, lessons about what works and does not work in diverse operating environments are drawn and disseminated. Accordingly, the evaluation methods used by OED make eclectic use of all major evaluation paradigms.

In brief, as the Bank has moved towards a comprehensive approach to development and a deliberate focus on partnerships as instruments of assistance, evaluation has shifted toward the lower right hand of the triangle (Figure 1). It continues to make room for self-evaluation and comprehensive auditing principles. With the help of Carol Weiss, OED has experimented with theory-based evaluation. Various forms of participatory evaluation have also been introduced and much of what OED now does is in line with the realistic evaluation principle of giving due weight to the 'context' when evaluating the development effectiveness of programs and processes.

Under its renewal program, OED has increased the professionalism of its work

Box 2. Mainstreaming through Evaluation Capacity Development

Bank loans and grants for evaluation capacity development have been approved or are under preparation for Brazil, Ghana, Guinea, Honduras, Niger, Poland, Romania and Uganda. Governments seeking Bank advice include Egypt, Lesotho, Malawi, Madagascar and Tanzania. Homegrown initiatives with potential outreach impact are found in Chile, Columbia, the Kyrgyz Republic and South Africa. Thus, the World Bank's approach to evaluation capacity development encompasses a broad array of tools and techniques adapted to country-specific needs.

Ten criteria guide the World Bank's mainstreaming of evaluation in its developing member countries:

(i) connection to public sector reform programs;
(ii) results-orientation and poverty-reduction focus;
(iii) part of public expenditures management improvement at central, sector and local levels;
(iv) involvement of the private sector and the civil society;
(v) co-ordination with other development assisstance agencies;
(vi) customized training;
(vii) linkage to financial management and accountability programs;
(viii) linkage to statistical improvement program;
(ix) linkage to development research and public policy analysis;
(x) contribution to improved monitoring and evaluation of poverty reduction strategies and externally funded projects and programs.

and the transparency of its activities. Starting in January 2002, all evaluation reports will be made public. Furthermore, OED is making more systematic use of the new information technologies and its outreach program of publications, workshops and conferences has expanded rapidly. As the largest evaluation group in the development assistance system, OED has become responsive to the broader needs of the development evaluation community. In particular, it is exercising leadership in evaluation harmonization and training. It has also given new impetus to its outreach activities by linking up with development 'mavens' (in the research community), 'connectors' (e.g. the World Bank Institute) and 'salesmen' (Country and Network Directors).

Together with the United Nations Development Program, OED is sponsoring the creation of an International Development Evaluation Association. Similarly, in partnership with like-minded evaluation units from other development assistance agencies it has put fresh emphasis on evaluation capacity development in developing countries. This is part of the World Bank's new major area of emphasis-governance defined as the nature, function, capacity and performance of the public sector.

Globally, evaluation capacity development is critical to evaluation mainstreaming. It is also congruent with the poverty reduction mission of the Bank: research has established that good governance is a primary determinant of sustainable and equitable growth. Key governance issues include:

- anti-corruption programs;
- civil service reform;
- improved public expenditures budgeting, auditing and management;
- participation of the civil society in determining national priorities; and
- encouragement of private investment through an appropriate regulatory environment, transparency, accountability and the rule of law.

Results-based performance monitoring and evaluation are critical ingredients of improved governance. These priorities lie at the core of OED's evaluation capacity development efforts (McKay, 2002) described in Box 2.

In sum, the World Bank has always sought to continuously develop and improve its evaluation capacity. The historical trajectory of OED has reflected the evolving needs of the institution as well as the priorities of the development community. The recent shift from project to program evaluation, the new emphasis placed on partnerships and the advent of evaluation capacity development as a major objective of development assistance have contributed to moving OED towards the mainstream of the evaluation profession just as major strides were made in moving OED activities towards the mainstream of World Bank business processes.

Notes

1. This article is a development of a presentation made by the author to the American Evaluation Association Conference in 2001.

References

Alkin, M. (2000) *Debates on Evaluation*. Newbury Park, CA: Sage.

Gladwell, M. (2000) *The Tipping Point, How Little Things Can Make a Big Difference*. New York and London: Little, Brown and Company.

Granovetter, M. (1973) 'Strength of Weak Ties', *American Journal of Sociology* 78(6): 1360–80.

Granovetter, M. and R. Swedberg (1992) *The Sociology of Economic Life*. Boulder, CO, San Francisco and Oxford: Westview Press.

Hirschman, A. O. (1970) *Exit, Voice and Loyalty, Responses to Declines in Firms, Organizations and States*. Cambridge and London: Harvard University Press.

MacKay, K. (2002) 'The World Bank's ECB Experience', in D. W. Compton, M. Baizerman and S. H. Stockdill (eds) *The Art, Craft, and Science of Evaluation Capacity Building, New Directions for Evaluation* 93: 81–99. San Francisco: Jossey-Bass.

Merton, R. K. (1996) *On Social Structure and Science*, edited by Piotr Sztompka. Chicago, IL: University of Chicago Press.

Milgrom, P. and J. Roberts (1992) *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice Hall.

Trilling, L. (1947) 'An Integral Part of Our Culture', *The American Scholar* 16(1).

World Bank (1992) *Natural Resource Management in Nepal* OED Precis No. 24. Washington, DC: Operations Evaluation Department.

World Bank (1993) *Borrower Ownership of Adjustment Programs and the Political Economy of Reform* Discussion Paper No. 199. Washington, DC: World Bank.

World Bank (1998) *India's Diary Revolution* OED Precis No. 168. Washington, DC: Operations Evaluation Department.

World Bank (1999) *Health Care in Brazil: Addressing Complexity* OED Precis No. 189. Washington, DC: Operations Evaluation Department.

World Bank (2000a) *Striking the Right Balance, World Bank Forest Strategy* OED Precis No. 203. Washington, DC: Operations Evaluation Department.

World Bank (2000b) *Agricultural Extension, The Kenya Experience* OED Precis No. 198. Washington, DC: Operations Evaluation Department.

World Bank (2001a) *Promoting Environmentally Sustainable Development* OED Precis No. 216. Washington, DC: Operations Evaluation Department.

World Bank (2001b) *Evaluating Gender and Development at the World Bank* OED Precis No. 200. Washington, DC: Operations Evaluation Department.

World Bank (2001c) *Chile's Model for Educating Poor Children* OED Precis No. 212. Washington, DC: Operations Evaluation Department.

World Bank (2001d) *Adapting Transportation Institutions to Romania's Transition Needs* OED Precis No. 207. Washington, DC: Operations Evaluation Department.

World Bank (2001e) *Community Forestry in Nepal* OED Precis No. 217. Washington, DC: Operations Evaluation Department.

ROBERT PICCIOTTO as Director General, Operations Evaluation, has led the independent evaluation function of the World Bank for the past ten years. His prior position was Vice President, Corporate Planning and Budgeting. He holds a Master's degree from Princeton University and an engineering degree from the Ecole Nationale Superieure de l'Aéronautique. Please address correspondence to: Operations Evaluation Department, The World Bank, Washington DC 20433, USA. [email: rpicciotto@worldbank.org]