

UKES ANNUAL CONFERENCE

Cardiff, December 1-3, 2003

Keynote Presentation: New evaluation standards for a new century

Robert Picciotto, Director, Global Policy Project

My purpose today is to relate the issue of standards to the overarching theme of this conference – the usefulness of evaluation. My message is threefold. First, standards are critical to the future of evaluation. Second, a results-oriented and comprehensive approach to standards should replace the current approach. Third, the new standards should be inclusive and universal.

These propositions seem far-fetched at a time when the momentum that propelled evaluation standards in the limelight during the 1990's appears to have waned. Certainly, the multiplicity of evaluation doctrines, the intellectual appeal of postmodernism and the dread of cultural imperialism have had a chilling effect. Yet, the issue of standards will not go away.

First, growing popular expectations and a generalized lack of trust in public institutions and private corporations are driving a booming demand for transparency, accountability and third party verification. Now more than ever, evaluators are facing a high and growing demand for their services. But paradoxically, there is also widespread unease about the uneven quality of evaluation products and the limited use of evaluation results. Generally accepted evaluation standards in combination with improved skills and sound evaluation governance would help strengthen the credibility of the profession.

Second, the center of gravity of the evaluation community has moved across the Atlantic and towards the developing world. The new public management is spreading to the zones of turmoil and transition where 80% of the world's population lives. A global evaluation community is in the making. It is seeking a common language and joint principles. The trend is irreversible since it responds to the same global forces that have led to major strides towards universal accounting standards.

Third, the positive role of evaluation standards has been demonstrated by their effective use in the development business. Evaluation is now an integral part of the governance of development assistance agencies. With development evaluation moving closer to the mainstream of the profession, the value of evaluation standards in real life and real time is becoming better understood.

Fourth, the vigorous debates evinced by standards within our profession are forcing a reconsideration of the traditional approach. New evaluation standards are needed for a new century. They should be results oriented. They should go beyond guidance to individual evaluators and address evaluation governance issues. They should be developed in a gradual, modular, comprehensive and inclusive way. Only then, will they secure broad based support within our expanding, feisty and free wheeling profession.

A Strong Prima Facie Case

At the origins of the evaluation business, the case for evaluation standards seemed rock solid –almost self-evident. After all, standards are what distinguish a profession from an occupation. All professions promote standards in order to achieve comparability and quality in the services they provide. By inducing self-control, standards generate public trust, reduce transaction costs and allow professionals (and the organizations that employ them) to operate with autonomy and efficiency.

On the supply side, well-designed standards provide quality benchmarks. They identify practices that enhance the validity, reliability, feasibility, propriety and utility of evaluations. They offer guidance regarding ethical issues. They also validate the need for governance safeguards (e.g. regarding public disclosure of evaluation reports). On the demand side, standards enhance the professional status of practitioners who commit themselves to operate in conformity with them. They facilitate comparisons among providers of services, thus helping customers secure value for money.

Tom Peters' famous phrase ("what gets measured gets done") highlights the role of incentives. Legitimate and broadly accepted standards are part of the enabling framework that encourages rigorous and objective professional work. Why should evaluation be the solitary exception to this precept? How without standards can the usefulness of evaluation be ascertained? For some proponents of utilization-based evaluation, the value of evaluation is measured as an integral part of the evaluation process—by the success or failure of the policies and programs being evaluated. This is not a sound practice: one does not judge auditors by the profitability of the companies they serve.

Evaluators are more than consultants. Their unique mandate is to assess the merit and worth of public policies and programs on behalf of citizens, consumers, the poor and the voiceless. While a byproduct of independent evaluation is to assist policy makers and program managers do a better job, the primary responsibility of the independent evaluator is to find the truth, verify it and tell it to power and the broader public. Just as accountants need auditors, evaluation consultants (and public officials who practice self evaluation) need independent evaluation.

Evaluators are accountable to the public – not to individual decision makers. The legitimacy of the profession is grounded in rigor, veracity and transparency. Recent corporate scandals in the United States have demonstrated the risks involved in breaching the firewall that separates external auditors from corporate managers. Similarly, asking the clients of evaluations to rule on their usefulness leads to moral hazard — i.e. "feel-good" evaluations tailored to what program owners and managers want to hear.

Do as I say – not as I do?

As a profession, we should practice what we preach. Standards occupy a privileged place in the theory and practice of evaluation. In order to "determine the merit, worth and value of things", the evaluator routinely identifies criteria of merit and performance measures (i.e. standards). Without tests of value adequately validated at the outset of an evaluation,

evaluators cannot assess results in a transparent fashion or reach a credible overall assessment of the success or failure of a public policy or program.

Carol Weiss refers to standards in her definition of evaluation as “the systematic assessment of the operations and/or the outcomes of a program or policy compared to a set of explicit or implicit standards.” Similarly, Michael Scriven has observed that: “evaluation has two arms, only one of which is engaged in data gathering. The other arm collects, clarifies and verifies relevant values and standards”. With implacable logic, he adds: “anything can be evaluated, including evaluation”.

Finally, according to Evert Vedung, “the value component of evaluation presupposes at least one criterion of merit against which public interventions are judged”. Since evaluation is a public intervention, it follows that the credibility of our profession requires generally agreed principles of professional conduct combined with quality norms for processes and products. If evaluators use standards to judge the performance of others, they should *ipso facto* recognize the need for standards as a basis for the assessment of their own work

Thus, far from being a peripheral issue, standards are central to the continued growth and usefulness of our profession. But if the case for evaluation standards is so strong, why has progress in setting evaluation standards been so slow, halting and partial? Why is the topic so controversial? Why is the resistance so widespread? This is the paradox that I wish to explore with you today.

Diverse Doctrines

An important obstacle lies in the diversity of evaluation doctrines. Different evaluation schools will elicit different criteria of merit. A pluralistic approach, sensitive to cultural differences is needed. We may well have to accept that universal agreement for certain types of standards (e.g. methodological or governance norms) will remain elusive. Hence, a modular approach to standards may have to be adopted. Beyond a central “core” of evaluation principles endorsed by all, each of the schools that co-exist under the big tent of our profession should adopt its own methodological standards. Such a differentiation will help users distinguish between the different evaluation specialties and branches that make up our profession.

For example, not all evaluators endorse the notion that evaluators have a mandate to judge the performance of public policies and programs¹. Some (e.g. Robert Stake) embrace the rich (or thick) description view of evaluation. They shy away from any evaluative conclusion – even from the standpoint of the program objectives themselves. Their “value free” doctrine reflects the traditions of anthropology and ethnography and they do not oppose standards. Indeed, they endorse precise rules of conduct with respect to the ethical collection and interpretation of evidence and the validity of findings.

¹ See Michael Scriven, Hard Won Lessons in Program Evaluation, New Directions Publication No 58

A variant – the social process school also rejects accountability and program management as proper concerns for evaluation. Its adherents view evaluation mostly as a tool for understanding social phenomena. But here too, standards (drawn from the canons of social science research) sustain the quality of evaluations. As early as the mid-fifties, the leading exponent of this tradition (Lee Cronbach) exercised intellectual leadership in the promotion of evaluation standards to help society guard against the corrupt practices of testing companies.

Similarly, the relativistic school of evaluation (represented by Rossi and Freeman) holds that evaluators are not expected to question the framework of values or objectives pursued by program managers and decision maker. Instead, the relativists focus on the discrepancies between what was intended and what actually happened. This approach has its place under the big tent of our profession. It resembles the methodology used by auditors who are utterly committed to professional standards of behavior and analysis.

The constructivists (Egon Guba, Yvonna Lincoln) and the empowerment evaluators (Fitterman) go further in their agnosticism. Their adherents reject any notion of objective benchmarks against which programs can be evaluated. At best, they view evaluation as a tool for facilitating the achievement of consensus among groups. But they too have standards about what good evaluation practices should look like as a tool of facilitation, participation or mediation.

In sum, standards are needed to protect the integrity and reputation of any profession. Here as elsewhere, cultural sensitivity is no excuse for professional mediocrity. Except for the deconstructionists that seek to infiltrate our ranks, no ontological rationale exists for rejecting evaluation standards as a matter of principle.

A question of semantics?

How then can we explain the widespread allergy to evaluation standards? Part of the explanation is semantic. The term is elastic and encompasses many meanings. Evaluation standards may address definitional issues (glossaries); professional ethics; analytical and rating methods; reporting processes or governance characteristics (e.g. evaluation mandate, independence, disclosure) etc. Indeed, the concept of standards needs broadening and unbundling all at once since different kinds of standards are needed to guide behavior in different circumstances and each raises distinctive challenges.

According to the Oxford English Dictionary a standard means “a rule, a means of judgment, a criterion”. But it can also signify “a thing serving as recognized example or principle to which others conform or should conform or by which the accuracy or quality of others is judged” or “an official statement of a rule or a creed” or even “a document embodying an official statement of a rule or rules” such as “a book or document accepted by a church as the authoritative statement of its creed” or “a document specifying nationally or internationally agreed properties...”

Some of these definitions have threatening connotations for evaluators committed to freedom of thought, diversity of perspectives and creativity in methods. The notion of

standards evokes the specter of standardization and its attendant risks: coerced uniformity, mindless orthodoxy, methodological rigidity and bureaucratic interference. Such negative perceptions are amplified when the original meaning of the term is evoked: “a flag or figurehead attached to the upper part of a pole and raised to indicate a rallying point”; the “distinctive ensign of a sovereign, commander, nation”; the “flag of a cavalry regiment as distinct from the colours of an infantry regiment, etc.”

Unsurprisingly, such religious and martial images elicit resistance among “free thinking” professionals who treasure the integrity of their craft and the freedom to select the evaluation methods most relevant to the specific evaluation challenges they face. Indeed, their concerns are not wholly misplaced. Evaluation standards improve the quality of evaluations only if they are judiciously targeted, designed and implemented in appropriate institutional environments. New standards for evaluation should be crystal clear about their limitations and address the broader institutional prerequisites without which evaluation results are unlikely to contribute to the welfare of society. This is the conclusion that emerges both from theory and from experience.

Insights from Theory

The two main functions of evaluation are to make authority responsible and to reduce the probability of errors in decision-making. These two roles (accountability and learning) are inextricably linked. They are two sides of the same coin². At the level of the individual decision maker, accountability for decisions taken provides incentives for learning while learning improves the quality of decisions.

According to Oscar Wilde, “experience is the name everyone gives to one’s mistakes”. But evaluation goes further: it helps not only individuals but also groups and organizations to learn from their own mistakes as well as the mistakes of others. This is far cheaper and less painful than trial and error. Thus, evaluation is a social function and its usefulness hinges on its performance along three distinctive dimensions: (a) collective action; (ii) participation; and (iii) knowledge creation.

First, evaluation contributes to the creation and implementation of a social or organizational consensus. Achieving shared objectives among group members is a prerequisite of effective *collective action*. It helps to secure the gains of specialization through *coordination* within the contractual framework that drives and constrains public choice. Since neither the state nor private agents have the power to unilaterally define their actions, social decision-making involves bargaining. Evaluation improves the social rewards of the game by providing impartial evidence to all parties and facilitating progress towards principled solutions. This is where *ethical* standards in evaluation come in. In particular, *propriety* standards ensure that evaluation is conducted with regard for the welfare of all those involved in the evaluation and affected by its results.

² The notion that one needs to trade off accountability for learning is mistaken. It reflects a common confusion between the distinctive roles of evaluation consultants (who face strict limits in their capacity to criticize the performance of their employers and the policies they pursue) and independent evaluators (who face no such constraints).

Second, evaluation levels the playing field of the collective action game by helping to increase the responsibility of authority. Through *participation*, it amplifies the influence of those who may not have direct access to decision makers – employees, clients, the poor, etc. By providing objective validation or censure of the actions taken by those in authority, it encourages the powerful to behave responsibly and fairly. This is where *governance* standards for evaluation structures and practices are critically needed to combat the frequent undermining of independent evaluation by vested interests. In this context, the role of the independent evaluator in the construction of results based management systems and other real time monitoring and evaluation processes (that are an integral part of public sector management) needs codification. Not enough attention has been given to this interface in traditional evaluation standards. This gap has been filled by the “*evaluability*” doctrine. It should have pride of place in the new evaluation standards.

Third, evaluation contributes to knowledge creation and the removal of information asymmetries that helps to reduce the probability of unnecessary errors. Access to relevant information is often too costly to be secured by individual actors. Interpretation of the evidence requires scarce professional skills. By dispassionately examining the record of past interventions and putting social science disciplines to work for the common welfare, evaluation helps to overcome the prejudices and self-serving opinions of decision makers while providing valuable information and advice that can be used by all partners to the enterprise to enhance the outcomes of public policies and programs. In this context, *methodological* standards provide comfort about the usefulness of evaluation results by addressing issues of validity, accuracy, context analysis and meta-evaluation.

The three roles are complementary but they are distinct and each requires norms. Some evaluations emphasize consensus making; others accountability; still others learning. Major evaluation exercises may embrace all three functions, either in equal measure or not. Hence, just as many evaluation doctrines have emerged to meet a variety of needs, a pluralistic approach to evaluation standards is needed to embrace the diversity of situations that evaluators face in the world of real life applications. Conversely, explicit consideration of the standards applicable to each evaluation assignment should be part of good evaluation practice.

The limits of standards

The first responsibility of the standard-setter is to observe the Hippocratic oath: “first, do no harm”. The economics of standards depend on weighing their benefits against their costs. Substantial harm can result from *wrong signals* about program outcomes. Good evaluation acts as a safety net to minimize this risk. Conversely, unprofessional methods can have perverse impacts by providing misleading information and unwarranted assurances. Thus, judging any social intervention (including evaluation) against the wrong goals can lead to erroneous findings and recommendations, e.g. when the goals are misguided, are too easy to achieve or are unrealistic. This highlights the need for care and circumspection in the design of evaluation standards as well as adequate flexibility in their use to minimize the risk of distorted incentives.

The utility of evaluation activities is dependent on the likelihood that evaluation findings will be utilized. In learning organizations, objective evaluations are utilized to improve the relevance and impact of interventions and, as a result, unlock enormous benefits at modest cost. Not so in poorly managed organizations where the evaluation process tends to be captured for the personal benefit of those in authority. This means that the effectiveness of evaluation standards is highly dependent on the operating environment.

In learning organizations, standards help to keep evaluation transactions low. In rigid, poorly managed, unaccountable organizations, evaluation involves tense interactions, “cover ups”, intimidation – so that standards can raise rather than reduce transaction costs. In turn, this means that, beyond *evaluation supply* factors (addressing the right issues, conducting evaluations at reasonable cost, ensuring fairness and professionalism, etc.) the economics of evaluation hinge on *demand* factors. And it is precisely on the demand side that a gap exists in traditional evaluation standards.

The gap needs to be filled. Evaluation standards should relate good practices to the institutional conditions under which evaluation actually takes place. This means that evaluation governance standards ought to be developed for various kinds of users. Organizational ethics are needed to avoid moral hazard and the punitive use of evaluation, i.e. for censuring low on the totem pole instead of enhancing corporate accountability and organizational learning thus discouraging responsible risk-taking and innovation. The new evaluation standards should also give pride of place to evaluability.

Thus, institutions more than individuals should be at the center of the new evaluation standards. This means that technical soundness of methodological standards (e.g. with respect to the randomized design of impact assessments) is only one feature of the challenge that lies ahead. To have a wholesome impact, evaluation must be embedded within a suitable governance framework that the new standards should help to define. Furthermore, the new standards should be “owned” by the individuals, groups and organizations that use them. Only then will they devote the resources and the skills needed to abide by the standards.

This is why the very process of formulating standards will act as a social learning mechanism within the evaluation profession. Experience from accounting and other professions also suggests that the usefulness of standards hinges in part on their actual use and their independent verification. Effective utilization of standards is facilitated by authoritative accumulation of evidence from adjudicated cases, especially those where the application of agreed standards has been contested. Finally, tacit norms tend to spread from organizations that are recognized as leaders in their field to the rest of the profession. Thus, sharing of good practice and training programs act as transmission belts between standards and their effective utilization.

Lessons from accounting and auditing

I have argued that standards must be viewed as “rules of the game”. Like other institutions, they constitute “a set of constraints that governs the behavioral relations among individuals or groups” (Jeffrey Nugent). The constraints must be meaningful but

not so strict as to paralyze action. At the macro level, evaluation standards help to improve performance assessments of public policies and programs and attest to the effectiveness of public expenditures and government services. At the micro level, they guide evaluation processes nested in within organizations where agents and their principals seek to achieve public or private goals. Thus, evaluation standards are to the public sector what auditing standards are to the private sector. Hence, the lessons gained in developing accounting and auditing standards have relevance for evaluation.

There is widespread agreement in the accounting and auditing profession that financial reporting standards combined with verification and enforcement are critical to the consistency and comparability of corporate accounts. The vigorous debate about financial reporting standards that still rages between the United States and European branches of the accounting profession centers on the design of standards – not on whether they are needed. Accountants and auditors know that judicious reporting standards help in the effective and smooth functioning of private markets. Conversely, they endorse self-regulation and government oversight of their activities.

The professional associations of accountants and auditors devote considerable resources to standard setting and certification. They endorse the concept of international accounting standards. Such standards are needed to deal with the problems faced by multinational companies that operate in diverse national jurisdictions. The preparation of reports based on different national principles is costly and raises problems of public perception, e.g. when the same facts lead to different judgments of profitability and other aspects of financial performance. Conversely, uniformity enhances credibility, comparability and efficiency in business transactions and facilitates cross border investments.

The same demands for cross border consistency and transparency are emerging in the public and voluntary sectors under the pressure of globalization. Accordingly, the lessons that have emerged in the process of generating universal accounting standards should be heeded:

- *The critical role of ownership*: for legitimacy, standards should gain broad acceptance by professional associations and public authorities at all levels and this in turn requires that they be transparent, enforceable and useful.
- *The imperative of judicious tradeoffs between uniformity and relevance*: the advantages of credibility, comparability and efficiency that uniformity delivers may conflict with the quality of the rules and their adaptability to different operating contexts. Principled compromises are needed and second best solutions adopted.
- *The need to strike a balance between broad guidelines and detailed prescriptive rules*: Diminishing returns tend to set in from efforts to make standards ever more precise and rigorous. Standards should be clear, broad and indicative rather than obscure, detailed and mandatory. They should be as simple as possible but not simpler. From this perspective, Alexander Hamilton's wise words about constitutions are well worth pondering: "(they) should consist only of general provisions: the reason is that they

- must necessarily be permanent and that they cannot calculate for the possible change of things”.
- *The importance of independence and competence in standard setting:* the standard-setting body should be representative independent and isolated from vested interests. It should have access to advisors and have the support of high quality staff. It should include users as well as suppliers of services. Members should be selected for their technical knowledge, experience and skills operating in their personal capacity rather than as advocates of any national, regional or functional interest.

The Foundations of Evaluation Standards

For evaluation just as for accounting and auditing, standards are part of the social context of the profession. But the evaluation profession differs from the accounting profession in that it must tailor its methods in relation to the unique features of individual evaluation assignments. This makes standards far harder to develop. But it does not mean that the challenge cannot be met. Most practitioners acknowledge that good and bad practices can be distinguished and welcome the judgment of their peers about the quality of their work. There is considerable evidence that individual evaluators value the formulation and publication of standards by their professional associations.

Thus, the American Evaluation Association was responding to a felt need when it developed *Guiding Principles for Evaluators* (1994) that specify basic criteria for the professional conduct of evaluations, including ethical standards. The principles are necessarily very general and cannot be relied upon to provide pointed advice in individual cases. But this does not detract from their usefulness when supplemented by case studies, training and guidance by experienced evaluation managers. Within their limits, they have provided the profession with a serviceable framework for learning and sharing of evaluation experience.

Similarly, the *American National Standards Institute* and the *Joint Committee (JC) on Standards for Educational Evaluation* have published standards for the conduct of program evaluations. The second edition (1994) goes beyond the education field on which the first edition concentrated. It deals with ethical, contractual and methodological aspects. The standards were generated through heated debate among leading exponents of different evaluation persuasions. They give equal weight to program improvements and accountability. They were developed in consultation with policy makers and users of evaluation. A third edition is under preparation. It will improve attention to cultural diversity and propose a new standard of cultural competence.

In 2002, both the Swiss Evaluation Society and the German published official standards inspired by the US Joint Committee’s work but they streamline, refine, adapt and supplement them (by adding a new standard related to the specification of evaluation objectives and removing the standard on conflict of interest). The UK Evaluation Society has settled for good practice guidelines while on-going work by the Australian Evaluation Society is expected to reach beyond the JC tradition. Building on the prior Australian *Guidelines for the Ethical Conduct of Evaluations*, the standards under preparation are

expected to emphasize issues of utilization and integration of evaluation withing the policy process.

The Case for International Standards

Finally, the context for the new evaluation standards is global. Already, six years ago, Eleanor Chelimsky and William Shadish wrote: *“evaluation is becoming international but in ways that go beyond previous conceptions of what international means. International is no longer used only to describe the efforts of particular evaluators in individual countries around the world ... Today, evaluation is becoming international in the sense of being at the same time more indigenous, more global and more transnational.”*

The evaluation profession has become more diverse and *indigenous*. It has had to adapt to the vastly different circumstances found around the world. But its dissemination has been influenced by the spread of democracy and the advent of new public management principles. Without the pressures of global economic integration and the revolution of rising popular expectations associated with the new information technologies, evaluation would not have trespassed across international borders. Localization has gone hand in hand with globalization.

Evaluation is also acquiring *transnational* features because it is tackling programs that extend beyond national borders. There is no greater challenge to humanity today that global poverty reduction. Evaluators are gradually turning their attention towards global issues given the injustice of a policy framework which results in almost half of the world’s population subsisting on less than \$2 a day (2.8 billion out of 6 billion people) and a fifth of them (1.2 billion) on less than \$1 a day. A crucial task for evaluation in the new century is to help fill the global public management gap.

Finally, evaluation has become more *global* in the sense that most economic and social developments in one part of the world now affect the rest of the world. According to Kofi Annan: *“Ours is a world in which no individual and no country exists 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.”*

Given the increased diversity and complexity of evaluation products and services associated with globalization, the discipline of evaluation has come under stress since meaningful evaluation standards have yet to be agreed across cultures and evaluation domains. But the time is ripe to seek convergence regarding the central objectives that we share as evaluation professionals and to build bridges across the various doctrines that give our unique profession its verve, resiliency and adaptability.

A New Kid on the Evaluation Block

Standards will help keep the profession together as it trespasses across national boundaries. Given the internationalization of evaluation, development evaluators,

previously isolated from the rest of the profession, are joining the mainstream. They represent a large professional group by any standard. Hundreds of professional evaluators currently work in bilateral and multilateral aid agencies and in civil society organizations concerned with economic and social development in poor countries. A vast literature that reflects their experience and concerns now exists. Through capacity building and training, the size of this group is growing rapidly within developing countries.

Development evaluators are now connected across national and regional boundaries, viz. the evaluation cooperation group of the multilateral development finance institutions; the evaluation working party of the Development Assistance Committee of the OECD and the interagency evaluation working group of the United Nations. IDEAS (the International Development Evaluation Association) represented for the first time at this UKES conference is a recent addition to the family of professional associations. It aims at systematic networking of development evaluators on a global basis.

In 2003, at the inaugural meeting of the International Organization for Cooperation in Evaluation (IOCE), headed by your former President, Elliot Stern, the heads of several national evaluation organizations from developing countries evinced interest in adapting the US standards to the cultural contexts and special requirements of their countries. This reflects the fact that evaluation standards have been a major preoccupation of the development evaluation community since its origins.

Development evaluators have taken to standards like ducks to water. They have contributed to the design of results based management systems, scorecards and reporting initiatives within their respective organizations. They have pioneered evaluation methods at the level of projects, programs and policies. They have endorsed common evaluation principles under the aegis of the OECD. A glossary of terms used in development evaluation and results based management has been published.

The evaluation cooperation group of the multilateral development banks has produced good practice standards for project evaluations in the public and the private sector. Evaluation governance has been addressed by the publication by the World Bank of voluntary standards for the independence of the evaluation function informed by the experience of multilateral finance institutions.

The World Bank Evaluation System

My own experience as Director General, Operations Evaluation at the World Bank Group confirms the usefulness of standards. With respect to evaluation governance, the World Bank makes a clear distinction between independent and self-evaluation. The former reports to the Board of Executive Directors (who represent the shareholding governments) while the latter reports to senior management. This distinction is enshrined in the mandate of the Director General, Operations Evaluation which empowers the

independent evaluation function to exercise oversight over self-evaluation products and practices; to attest to the validity of self-evaluation and to rate the performance of the organization based on short form evaluations of completion reports, long form evaluations for a sample of completed operations³ and evaluation studies that address operational processes, programs and policies.

The Operations Evaluation Department has developed a mission statement, values and product quality standards. A standard rating methodology guides the evaluation of development operations. Since development operations involve shared objectives, distinct accountabilities and reciprocal obligations among development partners, the rating standards distinguish between the outcomes of individual operations (measured in terms of relevance, efficacy, efficiency, sustainability and institutional development impact) and the attribution of these results to the distinct performance of government agencies, non-state actors and the World Bank itself. Similar methods are used to rate the performance of country programs and corporate policies. The Director General, Operations Evaluation rules on ratings that are contested by operating staff or governments.

A specialized committee of the Board of executive directors oversees the function, protects its integrity and independence and ensures that management takes due account of evaluation results in framing policies, programs and projects. To encourage effective utilization of evaluation findings, the executive directors review country assistance evaluations and policy evaluations at the same time as new country assistance strategies and new operational policies are being considered. Evaluation disclosure standards ensure transparency in reporting and evaluation is part and parcel of corporate management processes.

The Quality Assurance Group that reports to management carries out regular assessments of the quality of appraisal and supervision of development operations and produces reports that help in the management of the operational portfolio. Quality indicators consistent with the methodology pioneered by the independent evaluation units have been developed by management to track the quality of operations in real time. The success indicators of on-going and retrospective evaluations are fed in the corporate score card that tracks all aspects of organizational effectiveness as part of the corporate planning and budgeting process. Thus, the independent and self-evaluation systems are standards based, synergistic and embedded within the governance of the institution.

In brief, development evaluation as illustrated by the World Bank example (there are many others) has made good use of standards. They have helped to protect the independence and the professional excellence of its activities. They have influenced the quality of development operations funded by aid through systematic monitoring and reporting on results, regular review of operational policies and contributions to knowledge management and training. They have shaped policy making in such diverse fields as forestry, gender, resettlement, the environment, social funds, micro-enterprises and large dams.

³ The independent performance assessment ratio is currently set at 25 percent of completed operations.

What next for development evaluation?

The development evaluation profession is now poised for another major advance at the international level. Results orientation has been accepted as a fundamental principle of development cooperation and it has begun to influence civil society organizations and multinational companies conscious of their corporate social responsibilities.

All United Nations members have endorsed the Millennium development goals. They emphasize poverty reduction, environmental sustainability and social progress. Improved arrangements for aid coordination are being put in place under the aegis of poverty reduction strategies owned by governments of poor countries in consultation with the private sector and the civil society. Evaluation and monitoring systems managed by the United Nations Development Program and the international finance institutions form an integral part of this development strategy.

Ultimately, evaluation capacity will have to be constructed “on the ground” thus giving a further impetus to the internationalization of evaluation (and increasing the need for common standards) by expanding the frontier of professional evaluation to the developing world. The emerging development evaluation framework will go beyond aid to embrace the major transmission belts of globalization. It will examine policies that impact on international trade, migration, intellectual property, private finance and the environment.

This means that the focus of development evaluation once exclusively directed towards the south will gradually embrace the impact of the north on the south. The unit of account that had gradually moved from the project level to the higher plane of the country assistance strategy will have to move up once again -- to the global level. Only then will the thorny problem of attribution in performance assessments be resolved credibly.

In parallel, international evaluation alliances currently pursued through bilateral arrangements will have to evolve towards multilateral forms. This will make it possible to put the country at the center of country assistance evaluations and the global development community at the center of sector/thematic evaluations while avoiding evaluation bombardment. Finally, if development evaluation is to be judged by its results then the processes through which evaluations are conducted will have to “owned” by the partners and designed with the same amount of care as the methodologies.

The Road Ahead

It is time to conclude. I have argued that generally accepted standards are needed for our profession to demonstrate its usefulness. A comprehensive approach to standards (capturing its ethics, its governance, its methods and its linkages to policy making and resource allocation processes) will be needed so that the sterile debate between principles

based and rule based standards that has plagued the development of universal accounting standards is not repeated and the results based approach that is the hallmark of our profession is given a chance. The standards will have to be inclusive, embrace new stakeholders, accommodate all evaluation doctrines and focus on institutions rather than the individual evaluator.

First, the new evaluation standards should accommodate a worldwide trend towards market-based solutions as well as the advent of grassroots civil society organizations that are challenging traditional sources of authority and pressing for reforms. The new standards should also promote the objectives of the new public management movement that has brought results based management and evidence based policy making into government and induced development evaluation to join the mainstream of the profession.

Second, the design of evaluation standards will have to be inclusive by embracing all evaluation doctrines and delineating distinctive standards for each. Evaluation is a meta-discipline with antecedents in the social sciences as well as in the accounting and auditing profession. Extraordinarily diverse views will have to be reconciled through voluntary participation, principled dialogue and intellectual leadership.

Third, evaluation standards will have to be anchored within the institutional and cultural environment. Evaluation practice is necessarily influenced by the system of governance that evaluation serves and holds to account. Accordingly, evaluation standards will have to address evaluation governance and evaluability norms. Beyond the individual evaluator, standards should be directed to evaluation organizations and to the institutions that use evaluation.

Fourth, professional evaluation societies will have to take the lead in the design of evaluation standards. A gradual, organic progress is more likely to yield ownership than hasty standardization. This means that the design of evaluation standards should not be approached merely as a technical challenge. Just as war is too important to be left only to generals, the rules that govern evaluation cannot be left entirely in the hands of evaluators. To provide legitimacy and credibility to the formulation of evaluation standards, policy makers and representatives of the private and voluntary sectors should have their say and due process, including broad based public consultations, will have to be observed.

The challenge is ominous but substantial resources exist to carry out the task. The same forces that have propelled evaluation to prominence nationally, regionally and internationally must be mobilized. Undoubtedly, the voyage towards a universal, comprehensive and inclusive set of standards will be perilous and take a long time. But the potential rewards are great. So, let us start together, right away and right here in Cardiff.

Thank you for your attention.