

The Value of Evaluation Standards: A Comparative Assessment

Robert Picciotto

Following an exposition of the ethical dimension, professional role and democratic rationale of standards in the evaluation community, this paper proposes an assessment framework for rating evaluation standards, illustrates its use on a sample of published norms¹ and offers lessons for the participatory elaboration of global evaluation standards.

The Meaning of Standards

Dictionaries do not draw sharp distinctions between principles, guidelines and standards. According to the Oxford English Dictionary, a *principle* is a proposition serving as the foundation of belief or action; a guideline is a general rule or piece of advice; and a standard means 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.

Thus, the words tend to be used interchangeably although the notion of principles is commonly perceived as aspirational; guidelines are frequently intended as

¹ The sample reviewed in this paper includes Australia/New Zealand, Canada, France, Germany, Switzerland, the United States and the United Kingdom.

recommendations that do not take precedence over the judgment of experienced practitioners² while standards is the preferred term for mandatory norms, accompanied by enforcement or certification mechanisms.

Since this paper evaluates the intrinsic value of the norms rather than their application it makes no distinction between principles, guidelines or standards. In any event, since no enforcement or certification mechanism exists within the fledgling evaluation profession, all published evaluation principles, guidelines or standards are predicated on voluntary rather than mandatory compliance³ so that the difference between the terms is largely stylistic.

The Ethics of Standards

In industry, standards are used to impose uniformity in design characteristics or processes. They are needed to meet the demands of mass production and/or international commerce for goods and services. As a social practice on the other hand, standard making is designed to shape human behavior and interaction⁴. They

² For more precise definitions see: American Psychological Association, Board of Educational Affairs, *Developing and Evaluating Standards and Guidelines Related to Education and Training in Psychology, Context, Procedures, Criteria and Format*, Approved by the APA Council on February 20, 2004.

³ Principles and guidelines can be made mandatory by including them in contractual agreements between commissioners and evaluators.

⁴ Using the taxonomy of Marie-Louise Bemelmans-Videc, Ray C. Rist and Evert Vedung, *Carrots, Sticks & Sermons: Policy Instruments and their Evaluation*, Transaction Publishers, New Brunswick. 1998, guidelines are carrots, standards are sticks and principles are sermons.

help to achieve explicit or implicit policy goals. Intendedly or not, they promote the interests of particular groups and can restrain competition and creativity.

Hence, standard setting is legitimate only if provides for lack of coercion, equal treatment and the informed consent of participants in an open process. By clarifying expectations and setting rules of conduct, professional standards promote accountability, facilitate comparability and enhance the reliability and quality of services provided. They imply shared values, dedication to professional excellence and voluntary compliance with ethical guidelines. In democracies, standards are set in the public sphere and usually involve the civil society.

According to Jurgen Habermas, rational discourse among principled individuals is the only way to generate sound standards for knowledge creation: *“Representations and descriptions are never independent of standards. And the choice of these standards is based on attitudes that require critical consideration by means of arguments, because they cannot be either logically deduced or empirically demonstrated.”*⁵ This means that standards are context dependent and dependent on the outcome of deliberative processes that are shaped by specific cultural environments.

The Professional Dimension

Whatever their label, all existing evaluation norms have been socially constructed through rational deliberation and context dependent processes. No consensus has

⁵ Jurgen Habermas, *Knowledge and Human Interests*, Polity Press, 1968

yet been reached within the global evaluation profession as to the desirability of complying with internationally accepted norms. Thus, this paper is only meant as a contribution to an on-going debate about the future of the evaluation profession.

In most societies, principles, guidelines and standards are what distinguish a profession from a mere occupation. For some occupations, formal barriers to entry (e.g. academic degrees; certifications or licenses) help to protect the integrity of the profession. For others, informal criteria (e.g. a period of apprenticeship or a record of competitive achievement) suffice. But invariably the franchise enjoyed by a professional group is grounded on the presumption that its members are committed to live up to rules of conduct that protect the public interest⁶.

Such rules underlie the social contract that allows professionals (and the organizations that employ them) to enjoy public trust, practice their craft without undue interference and charge for services rendered. On the supply side, standards enhance the professional stature of those who operate in conformity with them and promote good practices. On the demand side, they facilitate comparisons among providers of services, thus helping customers secure value for money.

Even if the case for evaluation standards is accepted in principle, there are differences of views on their desirable range and scope. Evaluators are still debating whether it is appropriate to set uniform standards to guide or control how evaluation professionals, commissioners, participants and users should behave

⁶ According to the Wikipedia Encyclopedia, to conduct oneself as a professional is to act in accordance with specific rules, written or unwritten, pertaining to the standards of a profession. Evaluation being a young profession, it has yet to develop internationally agreed standards.

(ethical norms), what concepts and practices evaluators should use (methods), the benchmarks their products should meet (quality), the outcomes they should achieve (utilization) or the instruments needed to ensure that agreed standards are met and results achieved in the public interest (verification).

Standards as a Democratic Imperative

According to David Marquand⁷, democracy is characterized by a public domain where “citizens collectively define what the public interest is through struggle, argument, debate and negotiation.” Central to this process is an ethic of public service that “puts public duty and the public interest before market rewards and private interests”. In turn, the effective functioning of a pluralistic society requires professionals that operate autonomously, according to ethical standards. In such a system, professionals are motivated not by “patronage, clientism, connection” but by “professional pride, professional duty, professional authority and ...professional career paths”.

The mandate of evaluation is to assess the merit and worth of public policies and programs on behalf of citizens and with their participation and the credibility of evaluation hinges on proper assurances of quality, objectivity and independence.

Public trust is the critical ingredient. Absent certification and accreditation, evaluators must take it upon themselves to “internalize a set of norms precluding them from abusing their monopoly position and exploiting their clients, and enjoining them to promote the public good”.

⁷ David Marquand, *Decline of the Public*, Polity, 2004

Evaluation standards in and by themselves do not generate good policy outcomes. Indeed, they may have unintended consequences depending on how they are generated and used. If they are centrally imposed and coercively implemented, they may have a chilling effect on creativity and innovation. They may also be viewed as redundant where the value of the evaluation services provided can be reliably gauged in terms of the impact on the quality of decisions reached (ascertained as an integral part of the evaluation process).

On the other hand, just as one does not judge auditors by the profitability of the companies they serve, it is inappropriate to judge evaluators by the effectiveness of the programs and policies being evaluated. While a byproduct of independent evaluation is to assist policy makers and program managers do a better job (the advisory dimension of the profession), the primary responsibility of the evaluator in a democratic society is to enhance accountability, tell truth to power, illuminate policy options, promote public involvement and contribute to the transparency of decisions taken in the public interest.

Equally, asking the clients of evaluations to rule on their usefulness involves moral hazard. It may lead to “feel-good” evaluations tailored to what program managers want to hear rather than forthright assessments that protect the public interest. Evaluation needs to be responsive to the public interest and to operate without fear or favor. As for the accounting profession, the legitimacy of evaluations carried out by (or on behalf of) program managers requires independent validation against agreed standards.

In other words, for evaluators just as for accountants, the client is not always right. Evaluation needs to be independent of vested interests, including those of

sponsors⁸. Irrespective of the funding source, evaluators are ultimately accountable to the public.

While they should give close attention to issues raised by stakeholders, their credibility and integrity hinges on their objectivity and impartiality. This is why evaluation consultants as well as public officials that carry out self-evaluations need independent oversight that attests to the professionalism of their behavior and the validity of their findings.

Do as I Say – Not as I Do?

Ultimately, the case for evaluation standards rests on the golden rule: evaluators should practice what they preach. It would be perverse for evaluators to reject the discipline that they impose on others.

In order to determine the merit, worth and value of an activity, evaluators routinely identify relevant criteria of merit, i.e. they use standards to assess the results of programs and the performance of public officials. Thus, Carol Weiss⁹ 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.”

For Evert Vedung¹⁰, “the value component of evaluation presupposes at least one criterion of merit against which public interventions are judged”. In turn, Michael

⁸ This is why the notion of collective responsibility in evaluation is inappropriate.

⁹ Carol H. Weiss, *Evaluation, Second Edition*, Prentice Hall, Saddle River, 1998

¹⁰ Evert Vedung, *Public Policy and Program Evaluation*, Transaction Publishers, 1999

Scriven¹¹ observes 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”.

The design and adoption of commonly agreed evaluation standards would help to resolve the dilemma the profession currently faces in managing an exploding demand for evaluation services within an operating environment characterized by widespread unease about the uneven quality of evaluation products and services and the limited utilization of evaluation results. In setting standards, the emerging discipline of evaluation would emulate the experience of its forebears in the social sciences and the accounting profession.

But in order to make progress along the road of common evaluation standards, a paradox must be explored: if the case for evaluation standards is so strong, why has progress in formulating and endorsing evaluation standards been so slow, halting and partial? What are the limits of standards and what risks must be managed while designing and using them?

The Limits of Evaluation Standards

Evaluation determines the merit, worth and value of things¹². It consists in collecting relevant evidence, identifying suitable evaluative standards and using methods of analysis that are valid and fair. Assuming a stable and predictable

¹¹ Michael Scriven, *Evaluation Thesaurus: Fourth Edition*. Sage Publications. Newbury Park, London and New Delhi. 1991

¹² Michael Scriven, opus cit.

operating environment and provided the causal links that make up a results chain are known (and all evaluation actors are willing and able to comply with the reciprocal obligations that the agreed rules of the game imply) it should be enough to control the quality of outputs or to verify the ultimate impacts of an intervention to create the incentives needed to achieve the desired results.

In other words, the notion of standards is often associated with a linear conception of society predicated on rational behavior and predictable consequences of public policy actions. But in the real world, unintended consequences prevail, the operating context is unstable and the behavior of social actors is influenced by vested interests. The causal links between policy actions and impacts are subject to a wide range of external influences. Lack of precise knowledge regarding the connections between inputs, outputs, outcomes and impacts distorts decisions. This means that evaluation standards must take account of the volatility, complexity and contingency of public service tasks.

Like other social rules and protocols, standards are justified only if they generate social benefits in excess of their costs. Inappropriate standards can cause substantial harm by providing unwarranted assurances. Thus, crude performance indicators, simplistic league tables and performance assessments connected to inappropriate goals can destroy trust and erode the public service ethic. In particular, standards focused on only one stage of the results chain and rigidly applied (e.g. budget controls; quality assurance; inspection; auditing or evaluation) can create perverse incentives.

Conversely, applying and verifying standards at all stages of the results chain can lead to excessive rigidity and prohibitive transaction costs, especially where standards are mandatory and controls are tight. The dogmatic use of standards is

evoked by the alternative dictionary definition of standards as “a document accepted by a church as the authoritative statement of its creed”. Concerns about its restrictive implications can also be traced to the original meaning of the term: “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 colors of an infantry regiment, etc.”

Such martial images have threatening connotations for evaluators committed to freedom of thought, diversity of perspectives and creativity in methods. They evoke the specter of coerced uniformity, mindless orthodoxy, methodological rigidity and bureaucratic interference. Hence, the strong resistance to mandatory norms among “free thinking” professional evaluators who treasure the integrity of their craft and the freedom to select the methods most relevant to the evaluation challenges they face. This also explains the predilection of most professional associations for terms that are less threatening (i.e. charter, guidelines, principles, etc.) than the word standards.

Diverse Doctrines

The potential chilling effect of standards justifies a gradual and participatory approach to their design and adoption. A pluralistic approach, sensitive to cultural differences is fundamental. Special care is needed to avoid favoring one evaluation school over another. Not all evaluators endorse the notion that evaluators have a mandate to judge the performance of public policies and programs¹³. Some

¹³ See Michael Scriven, *Hard Won Lessons in Program Evaluation*, Sage, New Directions Publication No 58.

conceive of evaluation as a tool for understanding social phenomena. Others hold the view that evaluators are not entitled to question the framework of values or objectives pursued by program managers. Still others view evaluation as a tool for facilitating the achievement of consensus among groups.

Given this diversity, it is entirely legitimate for distinct evaluation schools to elicit different criteria of merit. This is why universal agreement for methodological norms has proven elusive. But all evaluation doctrines endorse judicious rules of conduct with respect to the ethical collection and interpretation of evidence, the validity of findings, etc. Thus, no ontological rationale exists for rejecting evaluation standards based on the notion that evaluation doctrines are manifold. On the other hand, due care should be taken to preserve the space that evaluation professionals need to practice their craft in line with their distinctive values and principles.

Beyond a central “core” of evaluation principles endorsed by all, each of the schools that make up the multi-faceted evaluation profession may choose produce its own principles and methods. Indeed, transparency about the methodological standards used in evaluations ought to be encouraged: clarity about the purposes and roles the evaluation methodologies is designed to serve would help users make a reasoned choice among evaluation suppliers, in line with the “truth in labeling” principle.

Lessons from Accounting and Auditing

Standards are “rules of the game”. They define roles as well as desirable outcomes. They set the voluntary restraints that govern the behavioral relations among individuals or groups. These must be meaningful but not so strict as to paralyze action or hinder innovation. They can be used to assess the performance of all

parties to an evaluation. Whether carried out by individuals or evaluation organizations, they guide the design of evaluation processes.

Since evaluation is to the public sector what accounting and auditing is to the private sector, the lessons gained in the process of developing accounting and auditing standards are instructive. In corporate finance, reporting standards combined with verification and enforcement guarantee consistency and comparability of accounts. The vigorous debate about financial reporting standards between the United States and Europe has centered on the design of standards – not on whether they are needed. Judicious accounting and auditing standards help in the effective and smooth functioning of private markets.

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 meant to overcome the problems faced by multinational companies that operate in diverse national jurisdictions. The preparation of reports based on different national principles undermines public trust in corporate accounts since different judgments of financial performance for the same multinational company resulted from inconsistencies in national accounting standards. Thus, the pressure for uniformity in accounting rules rose to protect the credibility, comparability and efficiency of business transactions and facilitate cross border investments.

Similarly, with the internationalization of evaluation under the pressures of globalization, national policies and programs can no longer be held to account without a clear set of benchmarks or without reference to their cross-border implications. A global evaluation community is emerging, peer reviews of national policy performance are becoming routine and a growing international consensus

behind new public management principles is fueling a demand for cross border consistency and transparency in public policy and program evaluations.

Thus, the lessons that have emerged in the process of generating universal accounting standards may have relevance for the design of evaluation standards¹⁴:

- **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.
- **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, where necessary, second best solutions should be adopted.
- **Need for restraint in prescriptive rules:** Diminishing returns result 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 simpler¹⁵.
- **Independence and competence:** the standard-setting body should be representative, independent and isolated from vested interests. It should have access to expert advisors and have the support of high quality staff. It should include users as well as suppliers of services. Members should be selected

¹⁴ John Flower with Gabi Ebbers, *Global Financial Reporting*, Palgrave, New York, 2002.

¹⁵ From this perspective, Alexander Hamilton's wise words about constitutions are relevant: "(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".

for their technical knowledge, experience and skills operating in their personal capacity rather than as advocates of any national, regional or functional interest.

The Genesis of National Evaluation Standards

For evaluation then, just as for accounting and auditing, standards are part of the social context of the profession. But to a far greater extent, the evaluation profession should adapt its methods to the unique features of individual evaluation assignments. This makes uniform standards for evaluation of public policies and programs far harder to develop than for accounting and auditing. On the other hand, most evaluation practitioners agree that good and bad practices can be distinguished. They accept the judgment of their peers about the quality of their work and they appreciate guidance about the conduct of their work.

In 1994, the Joint Committee on Standards for Educational Evaluation in the United States issued program evaluation standards¹⁶. Since then, other professional associations have generated their own guidelines, principles or standards. The formulation and publication of standards by professional associations has been welcomed even as their limits have been acknowledged. Thus, the American Evaluation Association was responding to a felt need when it developed *Guiding*

¹⁶ The Joint Committee was created in 1974. It published standards for evaluation of educational programs, projects and materials in 1981 and personnel evaluation standards in 1988. The Joint Committee was accredited by the American National Standards Institute (ANSI) to work on program evaluation standards in 1989. Student Evaluation Standards were published in 2003. The Joint Committee Program Evaluation Standards were approved by ANSI in 1994.

Principles for Evaluators (1994) that specify basic criteria for the professional and honorable conduct of evaluations¹⁷.

The principles are 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 evaluation profession with a serviceable framework for learning and sharing of experience about the ethical conduct of evaluations.

Similarly, the *Joint Committee (JC) on Standards for Educational Evaluation* published influential standards for the conduct of program evaluations. It deals with ethical, contractual and methodological aspects. The standards were generated through debate among leading exponents of different evaluation persuasions. They were framed in consultation with policy makers and users of evaluation in the education profession. A third edition is under preparation. It is expected to improve attention to cultural diversity issues.

The Swiss Evaluation Society, the German Evaluation Association and the African Evaluation Association have published official standards inspired by the US Joint Committee's work and subsequently streamlined, refined or adapted¹⁸. By contrast, the UK Evaluation Society's good practice guidelines issued in 2003 address

¹⁷ Peter H. Rossi, Howard E. Freeman, Mark W. Lipsey, *Evaluation: A Systematic Approach*, Sixth Edition, Sage Publishers. Thousand Oaks, 1999

¹⁸ The African Evaluation Association guidelines (not reviewed in this paper) include modifications in thirteen out of thirty US PES standards.

explicitly the distinctive roles of evaluation commissioners, evaluators and participants and they also provide guidance for institutional self-evaluations.

The concise charter issued by the French Evaluation Society in the same year stresses the commonality of obligations of evaluators and evaluands while ongoing work by the Australasian Evaluation Society is expected to reach well beyond the rights and responsibilities of commissioners, evaluators and other stakeholders codified in its 2002 *Guidelines for the Ethical Conduct of Evaluations* in order to address more explicitly issues of utilization and integration of evaluation within the policy process.

Criteria of Value for Evaluation Standards

Standard setting in evaluation needs to address social learning as well as individual learning. As Oscar Wilde famously observed, “experience is the name everyone gives to one’s mistakes”. But evaluation helps individuals, groups and organizations learn both from their own mistakes and from the mistakes of others. This is far cheaper and less painful than trial and error. At the level of the individual decision maker, accountability for decisions taken provides incentives for learning while learning improves the quality of decisions and broad based participation helps to protect the public interest¹⁹.

Accordingly, the three main functions of evaluation are (i) to reduce errors in decision-making through knowledge, (ii) to make authority responsible through

¹⁹ 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 are loath in their capacity to criticize the performance of their employers and the policies they pursue) and independent evaluators (who face no such constraints).

increased accountability, and (iii) to promote public involvement in public affairs. These three roles (accountability, learning and participation) are inextricably linked. Thus defined they help to determine how the profession should be judged. Specifically, evaluation standards should address three distinct dimensions: (a) collective decision-making; (ii) enhanced accountability; (iii) informed participation.

First, evaluation contributes to effective decision-making by nurturing principled solutions to complex public policy problems. 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 agreed solutions. This is where standards of *propriety* come in. They ensure that evaluation is conducted with regard for the welfare of all those involved in the evaluation and affected by its results. Resolving conflicts of value in a constructive way and achieving shared objectives among group members facilitates collective action. Through *transparency*, standards ensure that all parties and the public are adequately informed about evaluation findings.

Second, evaluation levels the playing field of the collective action game by helping to increase the responsibility of authority. By providing objective validation (or censure) of the actions taken by those in authority, it encourages the powerful to behave responsibly and fairly. Evaluation standards add value by discouraging the capture of the process by vested interests. By dispassionately examining the record of past interventions and putting social science disciplines to work, evaluation helps as a counterweight to the prejudices and self-serving opinions of decision makers. This is why *independence* and *integrity* standards are needed to help protect the value of evaluations.

Third, evaluation contributes to public involvement in decisions by reducing information asymmetries and providing advice to the public and to decision makers that can be used to enhance the outcomes of policies and programs. Through participation, evaluation amplifies the influence of those who may not have direct access to decision makers—employees, clients, the poor, etc. Access to relevant information is often too costly to be secured by individual actors. Effective collection and interpretation of the evidence requires *competence*. In this context, standards of *quality* and *comprehensiveness* of guidelines (to cover all stages of the evaluation cycle) provide comfort about the validity, accuracy and objectivity of evaluation findings.

Beyond the seven criteria implied by the three main functions of evaluation (propriety, transparency, independence, integrity, competence, quality and comprehensiveness), the net value added by evaluation is dependent on the *efficiency* of the evaluation process and the *utilization* of evaluation findings. These are only under the partial control of evaluators. Evaluation *governance* factors also intervene.

Getting Results

In learning organizations, evaluation helps to keep transactions low. In rigid, poorly managed, unaccountable organizations, evaluation leads to tense interactions, “cover ups” and even intimidation—so that evaluation ends up raising transaction costs with limited benefits in terms of improved organizational effectiveness. In learning organizations, objective evaluations are used 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, beyond evaluation supply factors (addressing the right issues, conducting evaluations with efficiency, ensuring fairness and professionalism, etc.) the economics of evaluation hinge on demand factors. How commissioners and program managers behave before, during and after the evaluation process is critical to the derivation of social benefits out of the evaluation process. This is why for evaluation standards to improve the functioning of society, it is important for evaluation standards to deal explicitly with the distinctive accountabilities and the reciprocal obligations of evaluators, evaluation commissioners and program managers.

In particular, the roles 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) would benefit from codification. Not enough attention has been given to this interface in traditional evaluation standards. Nor have the responsibilities of program managers been explicitly considered in the definition of evaluation standards. Conceptually and operationally, this gap has been filled by the “*evaluability*” doctrine²⁰. Looking ahead, it should have a place in evaluation standards geared to the achievement of results.

Ideally, evaluation standards should relate good practices to the institutional conditions under which evaluation actually takes place. This means that evaluation

²⁰ Joseph Wholey, *Handbook of Practical Program Evaluation* (Chapter 2), Jossey-Bass Publishers, San Francisco, 1994.

governance as well as evaluation practice would benefit from well-conceived standards. Without organizational ethics, the moral hazards of evaluation rise. In poorly managed organizations, evaluation is used punitively to name and shame. In such operating environments, managers use evaluation to censure those who are low on the totem pole. This is why evaluation standards should enhance accountability of authority as well as assist authority through learning from experience.

In other words, evaluation standards do not deliver optimum results unless they focus on the reciprocal obligations of all those involved in an evaluation. Institutions as well as individual evaluators should be guided by evaluation standards. Ethical considerations and technical soundness of quality standards matter but they should be embedded within suitable governance frameworks²¹. Furthermore, it is desirable that the standards be “owned” by the individuals, groups and organizations that use them. Only then are they likely to devote the resources and the skills needed to abide by the standards and make use of evaluation findings.

It follows that the very process of formulating and implementing standards should be viewed as a social learning mechanism (that is subject to evaluation).

²¹ According to Donald T. Campbell, “while all nations are engaged in trying out innovative reforms, none of them are yet organized to adequately evaluate the outcome of these innovations”. This observation led him to pioneer the concept of the “experimenting society” in which “policy relevant knowledge is created, critically assessed and communicated in real life or natural (not laboratory) settings, with the aim of discovering through policy experimentation, new forms of public action which signify a gain in the problem-solving capacities of society”. See William N. Dunn, Ed. *The Experimenting Society, Essays in Honor of Donald T. Campbell*, Policy Studies Review Annual, Volume 11, Transaction Publishers, New Brunswick, 1998

Experience from accounting and other professions suggests that the usefulness of standards hinges on their actual use and, in due course, 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. This is one reason why professional associations have a comparative advantage in the formulation and verification of standards.

A Comparative Assessment

A cursory comparison between national evaluation standards brings out the following common features:

- **Brevity:** the standards in the sample are invariably stated in concise and non-technical terms; they contain 3-6 principles and 11-44 guidelines. The listing of standards varies between one and six pages²².
- **Scope:** Most guidelines focus on the ethical conduct of public program and policy evaluations while the UK product also addresses institutional self evaluation standards.
- **Multiple audience:** all standards in the sample are directed to the individuals and organizations that commission, prepare, conduct and use

²² Additional space is often devoted to commentaries about the guidelines.

evaluations as well as to stakeholders affected by the evaluation or who have an interest in the results.

- **Process orientation:** the standards tend to eschew methodological aspects; instead they concentrate on behavioral, contractual and ethical considerations.
- **Asymmetry:** most standards give far greater weight to the responsibilities of evaluators and the rights of other stakeholders than to the obligations of evaluation commissioners and program managers with the notable exception of the UK guidelines.

This said there are substantial differences among the published standards. In order to carry out a comparative assessment among them, based on the considerations elaborated in the above sections of the paper, the ten criteria of merit that were identified above were used by the author as an evaluation template:

- **Propriety:** preservation of the dignity, security and privacy of people; protection of confidential information; prior informed consent of participants.
- **Transparency:** agreed expectations about objectives and methods are disseminated to stakeholders; evaluation reports disclosed to stakeholders and the general public without modification.
- **Independence:** adequate safeguards provided to ensure that vested interests do not influence the evaluation; distinct accountabilities are attributed to evaluation commissioners, program managers and evaluators; full protection is provided against intimidation and interference; adequate funding without strings is provided.

- **Integrity:** disclosure and avoidance of actual or potential conflicts of interest; contestability of evaluation judgments; evaluator’s access to relevant information.
- **Competence:** requirements about the knowledge, skills and experience expected of evaluators.
- **Quality:** guidance about the practices needed to achieve evaluation relevance, credibility, rigor and objectivity; norms for achieving fair and valid evaluation findings and recommendations; practices that generate constructive relationships among participants.
- **Comprehensiveness:** coverage of all phases of the evaluation cycle – from commissioning to dissemination and utilization.
- **Efficiency:** economy in use of skills, funds and administrative resources; limited burden on participants.
- **Utilization:** responsiveness to the public interest and to users’ needs; participation of stakeholders in the evaluation; timely delivery; clarity of presentation.
- **Governance:** clarity of roles between evaluation commissioners, evaluators and participants; appropriate contractual relationships; oversight of self-evaluation by independent evaluation; “evaluability” norms for program and policy design.

Table 1 presents the summary results based on the admittedly subjective judgment of the author. Equal weights were ascribed to each category. Based on a textual analysis of their content, each of the national standards was rated from 1 to 6—where 1 presents a highly unsatisfactory treatment and 6 a highly satisfactory

treatment. The exercise was carried out for illustrative purposes only, i.e. to demonstrate that evaluation standards can be valued. No claim is made for their accuracy. And it goes without saying that the process followed does not comply with sound evaluation process norms. Validation of the criteria by an expert panel combined with independent scoring by qualified evaluators would be required to attest to the reliability of the individual ratings.

Table 1. Ratings of National Evaluation Standards

	Australia	Canada	France	Germany	Switz.	UK	USA	Average
Propriety	6	5	5	4	5	6	5	5.1
Transparency	4	4	4	4	4	6	4	4.3
Independence	3	1	3	1	1	4	1	2.0
Integrity	2	2	2	2	4	4	2	2.6
Competence	5	5	5	5	5	5	5	5.0
Quality	5	3	3	6	6	4	6	4.7
Comprehensiveness	6	2	3	5	5	5	5	4.4
Efficiency	1	4	1	5	5	1	5	3.1
Utilization	2	2	2	5	5	6	5	3.8
Governance	4	3	3	3	3	4	3	3.3
Average	3.8	3.1	3.1	4.0	4.3	4.5	4.1	

Most national standards give considerable emphasis to the imperative of doing no harm and stress the rights of evaluation participants and the protection of confidentiality. Some mandate a right of prior informed consent for evaluation participants preferably in writing. In general, the standards give salience to the necessity of ensuring that evaluators have appropriate knowledge and skills. The critical importance of quality standards is also stressed, except for the charters of Canada and France that treat this aspect very lightly.

The lowest ratings are those related to the independence and integrity of the function. Only the United Kingdom achieves a satisfactory rating for independence given the scrupulous balance it strikes between the reciprocal obligations of evaluation commissioners and evaluators in achieving good evaluation outcomes. The Australasian guidelines and the French charter come next—the former because of its strong focus on fair and precise contractual relationships and the latter because it states unambiguously that arm’s length relationships between evaluators and program managers (“distanciation”) are needed to ensure the credibility of the process.

None of the other national standards address the risks inevitably associated with cases where evaluation commissioners have major executive responsibilities for the programs being evaluated. Integrity ratings are relatively low because conflict of interest problems tend to be treated lightly if at all and contestability procedures are not provided for. Where conflicts of interest are treated it is in terms of requiring their disclosure rather than on automatic disqualification from taking part in the evaluation.

Nevertheless, Switzerland and the UK achieve satisfactory ratings in this category, the former because of its emphasis on neutral reporting, the latter because it makes clear that the commissioners have a responsibility to provide evaluators with access to relevant documentation and data. The lack of reciprocity in the obligations of evaluation commissioners and evaluators is especially striking with respect to information disclosure. Evaluators are invariably instructed to make transparent the evaluative information on which they have based their findings.

On the other hand, the unimpeded access to relevant information (an acid test of independence for evaluators) while encouraged in some cases is not guaranteed by

any of the national standards. Nor under the rubric of transparency does the inalienable right of the public to access uncensored evaluation reports figure explicitly in any of the national standards although the UK guidelines discourage the quotation of evaluation results out of context and suggest that the final reports should “normally” be lodged in the public domain.

The Canadian guidelines do not address the disclosure of evaluation reports. Instead they emphasize the responsibility of evaluators to their clients with respect to confidentiality, privacy and ownership of findings and reports. The US guidelines (along with the German and Swiss versions that they have inspired) take a middle and somewhat ambiguous course by requiring that the “results” of the evaluation be made available to all potentially affected persons as well as to all others who have a legitimate claim to receive them.

Similarly, the French charter opines that public dissemination of evaluation results is desirable but reserves actual disclosure modalities to a negotiated outcome at the time of evaluation commissioning. By contrast, the Australasian guidelines are explicit in requiring the consent of the evaluator for any amendments to the final report but they do not compel the commissioners to disclose evaluation reports to the public. Instead, they enjoin commissioners not to breach the integrity of the reports in their pronouncements.

The Road Ahead

The above analysis brings out the following conclusions that may help trace a road map for future work on evaluation standards. Utilization ranks highest for the UK guidelines because they devote a full section to self-evaluation. Concern with utilization is also high in the US standards as well as the German and Swiss standards that they have inspired. Nevertheless, most of the evaluation standards

are not oriented to results. They give considerable weight to the contracting phase of the evaluation process, a stage when the commissioners have enormous leverage over the evaluator. They stress early identification of risks and promote good communications between evaluators and commissioners without specific provision for contestability, arbitration mechanisms or independent oversight of the executive branch by the legislature.

Most national standards give far more weight to the obligations of evaluators than to policy makers, program managers and evaluation commissioners. They do not address criteria of program “evaluability” or the measures needed to ensure effective utilization of evaluation results. They do not instruct evaluation commissioners to support evaluators in their evaluative work; provide them with unencumbered access to data; protect their independence; and avoid retribution, intimidation and other means of evaluation capture.

None of the standards makes public officials accountable for the effective use of evaluation results in the public interest. This would require the formulation of standards that address explicitly the institutional prerequisites of organizational learning, e.g. based on the accountability principles of the new public management movement. This would involve codification of the distinctive roles of independent evaluation, self-evaluation, inspection and auditing in various administrative environments.

For evaluation standards to be fully relevant, evaluators, evaluation commissioners and program managers would need specific guidance with respect to results based management systems, quality assurance processes, results based scorecards and selection and use of performance indicators in public service delivery. This is where the current frontier of program evaluation activities lies. Similarly, the

regulatory interface between citizens, government, voluntary organizations and the private sector would benefit from judicious guidelines. Other promising areas for standard setting include the design of appropriate linkages between independent evaluation, budget allocation processes and personnel evaluation practices.

National audit offices have often trespassed profitably into the evaluation domain through value for money and comprehensive audits. Conversely, systematic evaluations of the effectiveness of public auditing and inspection processes would have considerable merit and should be encouraged. In most industrial democracies, public officials feel victimized by “inspection overload”, taxpayers feel powerless to influence the quality of public services and performance indicators are widely criticized because they fail to encourage beneficiary involvement and genuine quality assurance. This suggests a need for more systematic evaluations of control functions, including of inspection and auditing...and of evaluation itself.

Towards Global Standards

Finally future work on evaluation standards should take account of the *transnational* features that now characterize the profession. Increasingly, evaluators are called upon to assess public policies and programs that extend beyond national borders. As a result, a global evaluation community is in the making. It is seeking a common language in order to facilitate evaluation assignments across national boundaries. Given this new context, harmonization of evaluation standards across national boundaries would be desirable. Demands for cross border consistency and transparency in evaluation have become more pressing.

But a global approach not grounded in national and regional experience would involve risks of coercion, rigidity and misplaced homogeneity. To achieve

credibility and legitimacy, global standards should be grounded in the initiatives of national evaluation associations. Consistent evaluation standards that would respect universal ideals of peace and justice would promote high quality work in evaluation, foster trust in the profession and contribute to the advent of an international evaluation community. There is now a wealth of experience in the design and implementation of national standards. It would be appropriate for such standards to be used as building blocks for a global initiative.

The global standards should be inclusive, embrace new stakeholders, accommodate all evaluation doctrines and focus on institutions rather than the individual evaluator. A comprehensive approach to standards (capturing its ethics, its governance, its methods and its linkages to policy making and resource allocation processes) would be desirable 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 the evaluation profession is given a chance.

National evaluation societies should take the lead in the design of global evaluation standards. A gradual, organic progress is more likely to yield greater ownership than hasty standardization. To provide 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. Last but not least, in order to ensure legitimacy, special efforts should be made to involve evaluators of the developing world where 85% of the world's peoples live.

About the Author

Robert Picciotto is Visiting Professor at King's College, University of London. He served as Director General, Operations Evaluation in the World Bank Group from 1992 to 2002.