

An Association to Improve Evaluation of Development Aid

Paul Clements

Western Michigan University

This paper argues that development aid could be greatly improved if the management of development programs and projects could be governed by an effective orientation to cost-effectiveness. This in turn could substantially be achieved by changes the paper proposes to aid evaluation—through a particular kind of association—and a few lesser management reforms. During the 60 years of development assistance, evaluation has largely been controlled by donor and implementing agencies. When it comes to the evaluation of their own programs, however, these agencies face profoundly mixed incentives. Moreover, due to the particular needs for learning and accountability in this field, the demands placed on evaluation in development aid are uncommonly great. When we think about how development interventions should be evaluated, we often start from the perspective of an individual program or project or from that of a single donor or implementing agency. This, I argue, is a mistake. We should approach aid evaluation from the perspective of the development assistance community as a whole. Given that the challenges of development are common challenges, both in the sense of being similar for different agencies and in the sense of arising from shared goals, we should look for evaluation approaches that best support the development community overall.

We can gain some insight into the current situation in aid evaluation by considering a typical low-income country with 20

development programs and projects in agriculture. Half to two thirds of these projects have some kind of evaluation. However the evaluation framework is worked out independently for each project, and the evaluators have a great diversity of academic training and practical experience. Evaluation criteria and approaches are inconsistent, and only one or two of the evaluations make reference to evaluations of other agricultural projects. Also, officials of the very agencies that funded and managed the projects select the evaluators. Since a negative evaluation is likely to hurt the professional and organizational interests of the evaluation commissioners, the evaluators suspect that they will be better off if they give positive findings. Few of the projects have serious monitoring systems, and it would be hard to estimate impacts. Except where things are really bad the evaluators report positive and negative findings regarding process and outcomes and partial achievement of goals. They conclude that overall the project was satisfactory.

This basic scenario, with minor variations, is played out repeatedly in most sectors of most low-income countries that have significant aid operations. How would this scenario change if the association this paper proposes were established? In this new scenario too probably about two thirds of the 20 projects would be evaluated, but here all twelve or thirteen evaluations would be based on common criteria of cost-effectiveness. The evaluators would

know evaluation, they would know the literature on approaches to the agricultural challenges facing the country, and they would share common conventions for estimating each project's total impacts and cost-effectiveness. They would be familiar with several evaluations of similar projects in this and other countries, and their assessments would be explicitly comparative, identifying the merits of the evaluated projects relative to other actual and/or possible projects. The shared conventions for estimating project impacts would not only render the estimates technically sound – they would also protect the evaluators, and hence the evaluations, from actual or perceived pressure for positive bias, as significant violations could lead to loss of membership in the association. Each evaluation would explain how the project's impacts (of low, medium or high cost-effectiveness) were achieved, beginning with an assessment of the project plan, and they would identify areas of particular strength or weakness in project design and management. Compared to today's evaluations, their analyses of project dynamics would be (on average) both deeper and more reliable, providing a sounder basis for learning and for accountability. Moreover, their recommendations would be better informed and more helpful.

If this scenario could be borne out, it would require those planning and managing projects to adopt an orientation to cost-effectiveness and it would support them in doing so. By reading evaluations and evaluation syntheses, managers would gain an integrated perspective on the factors that contribute to greater cost-effectiveness in their sector, and now they would have a stronger incentive to apply these lessons. This would lead to a reduction in major design errors, the establishment of better monitoring systems, and more rapid mid-course corrections in project designs. Better designs and design features would more rapidly be identified and replicated, and professional recognition and advancement would come to be

based more significantly on contributions to impacts. After a substantial initial investment routine evaluation costs might rise by perhaps 50% compared to today, but it is reasonable to expect gains in impacts to greatly outweigh increased costs.

To substantiate this argument, the paper proceeds as follows. The next section discusses the great challenges of learning and accountability in foreign aid and how the incentive environment under which evaluation has evolved has undermined evaluation's contribution to aid management. The third section lays out the design of the proposed association and discusses how it might operate, and the fourth section discusses its likely impacts. The fifth section deals with what it would take to establish the association, and the sixth section concludes.

Challenges of Learning and Accountability in Aid and Incentives Influencing Aid Evaluation

It is important to appreciate the unusual burden that evaluation bears in development aid. Compared to the tasks most organizations undertake, improving economic and social conditions in developing countries is particularly difficult. Standards of efficiency and effectiveness in public and private institutions in these countries tend to be low and working conditions difficult. Access to aid resources is carefully cultivated, and aid programs become enmeshed in political agendas and personal enmities. For most programs to succeed many people have to undertake new responsibilities and carry them out well, and coordination and logistics are typically quite demanding. Most development agencies work in many countries, but a strategy that works in one place may not work well in another. The history of development assistance is less one of building on and refining strengths than one of sequential

fads, with the new one displacing the former after it is found inadequate. We only hope that today's poverty reduction strategy papers and sector-wide approaches will not follow the pattern of state-led industrialization in the 1960s, integrated rural development in the 1970s, structural adjustment in the 1980s, and decentralization in the 1990s.

Development assistance routinely sets idealists to planning programs to be carried out by other people in a country that is not their own. Also, like evaluators, development planners and managers come with a remarkable variety of academic and professional backgrounds and ideological commitments. The competition for resources requires them to act as policy entrepreneurs, promoting the positive merits of the approach their program takes. All too often, however, the hard realities on the ground do not yield to planners' and managers' ideals.

For most organizations in other fields, even if the leaders' initial ideas are misguided, operational experience forces them to adapt to clients' needs and desires. In the private sector this works through the market mechanism based on customers' willingness to pay. For most organizations in the public sector, at least in a democracy like the United States, there are multiple channels through the political process for clients to express their views. Especially when services get really bad, representative governments are likely to hear about and to address the problem. For development agencies, however, it is a constituency in the home country that allocates resources that are later spent thousands of miles away. The intended clients have no role in the process. To sustain the flow of resources it is the home constituency that the development manager has to satisfy, but when a program is failing its clients have no recourse.

For all of these reasons, the challenge of learning to identify program designs and management strategies that are more likely to succeed in a given development context is

particularly great. At the same time, development planners and managers are less likely to be held accountable for the consequences of their choices than professionals in most arenas. We have already noted that development aid lacks the structural accountability of the market or the electorate. In addition, the consequences of management choices normally play out over periods of many years, and many other factors besides the project are sure to influence client conditions. Within the standard organizational framework of development management, the only way project results would normally come to be known is through evaluations.

Unfortunately, however, it is the very donor and implementing agencies that evaluations are supposed to assess that have controlled the evaluation process. Insofar as new funding depends on likely results, donor and implementing agencies have an incentive to make their impacts appear positive. Yet they organize the evaluations that provide the main evidence of their impacts. There are two ways that the evaluation process could be manipulated: by introducing or encouraging positive bias in evaluations, or by basing evaluations on less rigorous criteria than expected impacts and cost-effectiveness. For example, evaluations based on how a project carries out its main tasks (process) or on how it achieves its objectives (outcomes) often reach more positive findings than evaluations based on impacts and cost-effectiveness.

As prevailing incentives would lead us to expect, evaluation standards in international development aid have indeed been mixed and generally unsatisfactory. In fact for several decades the leading development agencies, first the US Agency for International Development (USAID) and then the World Bank, did evaluate many of their projects in terms of economic rates of return (ERRs), a measure of cost-effectiveness. However they never developed institutional means to render ERR estimates consistent between projects, and they are

increasingly moving to other evaluation approaches (e.g., Kusek & Rist, 2004). Development agencies generally have not made consistency between evaluations (in their evaluative judgments) a priority, and the selection of evaluators, particularly for the key end-of-project evaluations, has usually been decentralized.

There is only limited direct evidence of positive bias in aid evaluations. This author found overwhelming positive bias in evaluations from a blind selection of four World Bank and four USAID projects in Africa (Clements, 1999), and many authors have remarked on the gap between evaluation averages that are always positive and country-level trends of economic stagnation or decline. There is more consensus, however, on the poor quality of typical aid evaluations. For example, Robert Piccioto, a former Director-General of Evaluation at the World Bank, asserts that the record of international evaluation has been “dismal” (Bollen, et al., 2005, p. 190). Chianca finds that only about a quarter of US-based international nongovernmental organizations (NGOs) have developed evaluation policies and standards and less than 10% subject their evaluations to metaevaluations or carry out any regular synthesis of evaluation findings (2007, p. 129). In a study of over two dozen evaluations of USAID democracy and governance programs, Bollen et al. find “a lack of methodological accuracy and inappropriate coverage of important information about the impact of assistance interventions” (2005, p. 199). They conclude that their assessment:

...appears to match the assessments of other researchers who have reviewed the evaluations of interventions by other international development agencies. The typical evaluation lacks an appropriate research design, measures of inputs and outputs, and controls for confounding variables to justify sound assessments of whether an intervention accomplished its goals (*ibid.*, p. 202).

A review of World Bank agricultural projects in Africa from 1991 through 2006 similarly finds that:

M&E [monitoring and evaluation] at the project level has been of limited value in answering fundamental questions about outcome, impact, and efficiency, such as who benefited, which crops received support and how, what has been the comparative cost effectiveness, and to what can one attribute gains (World Bank, 2007, p. xxviii).

While there are many excellent individual evaluations, most professionals who deal regularly with aid evaluation would probably agree that the general standard is inconsistent and often poor.

In light of the organization of aid evaluation and the incentives bearing on its conduct, it would be surprising if standards were much higher. Compared to the situation in earlier decades, today’s development assistance community is more aware of the need to improve evaluation. Reforms to date, however, have been piecemeal. None has addressed the basic incentive structure or provided a means to achieve consistency between evaluations, the qualifications of evaluators remain extremely diverse, and the proportion of evaluations that estimate cost-effectiveness is probably declining. Unless these problems are addressed, however, the potential contribution of evaluation to accountability and to learning within the development community will not be realized.

Evaluation International

To reiterate, evaluations would do much more to serve the unusually great need for learning and accountability in the international development community if they routinely estimated the impacts and cost-effectiveness of the subject program or project in terms of criteria and on the basis of judgments that were technically sound and also consistent between

evaluations. This could be achieved in the same manner that accounts and audits for public corporations have been rendered (in most cases) technically sound, independent, and consistent: through the establishment of an association that controls entry to the profession and sets and enforces the standards. For the sake of discussion, let's call it Evaluation International (EI).

Evaluation International would address the problem of positive bias by establishing a general evaluation approach and principles and conventions for implementing it. The approach would be to estimate the impacts and cost-effectiveness of each program or project under evaluation and to explain how they were reached. It is essential that these should be estimated because impacts and cost-effectiveness provide the uniquely appropriate basis for learning and accountability in the international development community. The development community should do the best it can to improve conditions for poor people, and this means using its resources as cost-effectively as possible. All other criteria with the exception of serious ethical violations should be channeled through the prism of cost-effectiveness.

In this conception, impacts are understood as total changes in conditions of "impactees" compared to the conditions one would expect if the intervention had not taken place. When a project was begun with an evaluation system involving an experimental or quasi-experimental design, estimation of the counterfactual up to the time of the evaluation may be relatively straightforward. For most interventions, however, impact estimates involve an (implicit or explicit) reasoned assessment of the likely counterfactual. Also, total impacts attributable to a program typically include many that have not yet taken place, but estimates of future impacts are particularly difficult. Impacts include all changes in conditions attributable to an intervention, for intended beneficiaries and for others, that are relevant to any stakeholder. They may include improvements in income and

health, social capital, education and/or empowerment, as well as any harms (negative impacts) an intervention may have caused. In most cases the estimation of impacts is not merely a matter of calculation, but also one of reasoned assessment of evidence. Once all impacts have been estimated and costs tabulated the assessment of cost-effectiveness is a straightforward mathematical operation. Establishing principles and conventions for all of these assessments is a major part of establishing EI.

While cost-effectiveness orients the methodology, the EI approach would also incorporate standard principles of good evaluation practice, as expressed, for example, in Scriven's *Key Evaluation Checklist*.¹ Conventions are needed in particular for estimating future impacts in evaluations at project completion. They would indicate how different kinds of evidence, such as of institutional strengths and weakness, should be factored into impact estimates. Once the approach, principles, and conventions are worked out, the problem of positive bias is addressed by requiring adherence to them in order to retain membership in the association. This allows an evaluator to defend herself, in case she feels pressured to over-state a project's impacts, by noting the principle or convention that she is required to apply.

To implement this model, EI would need a guidebook to lay out its approach, principles and conventions, an examination to establish basic mastery (and hence admission to the association), and a standards committee to judge adherence. Presumably these three core components and their basic operations would be worked out in tandem. It would be reasonable to begin with a focus on three or four sectors, such as agricultural extension, primary health care, water, and microfinance. Then the association could use evaluations in these sectors to work out its conventions. As

¹ http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf

well as the guidebook, EI would also establish an online, indexed archive of evaluations with commentary on their strong and weak points. Each completed evaluation by a member of the association would be added to the archive, and the archive would be open to the development community at large. The archive would not only illustrate the application of the guidebook; it would also help managers to learn from one another's experience. Someone starting a new water project, for example, might consult evaluations of similar completed projects including some that were cost-effective and others with limited impacts.

When EI awards membership it would give each newly certified evaluator an EI stamp. When an evaluator stamps an evaluation she has completed, this indicates that in her professional judgment it is consistent with the EI approach, principles and conventions. The evaluator is obliged to make her best estimate of the evaluated program or project's likely total impacts and cost-effectiveness. Of course many interventions do not conduct baseline surveys or develop reliable monitoring systems, so there may be limited bases for estimating impacts at their completion. After conducting an appropriate survey and exploring exogenous influences on client conditions, the evaluator may only be able to estimate impacts quite roughly. Besides information from the project, however, she normally will also know evaluations of several similar projects and their estimated impacts and cost-effectiveness. These evaluations may indicate how characteristics of the project and its implementation are likely to contribute to impacts. The evaluator's task is to locate the present project within the range of evaluated projects based on its likely cost-effectiveness (loosely, impacts minus costs all adjusted for time). Supposing there are five cost-effectiveness categories (e.g. very low, low, medium, high and very high), she may well be able to exclude three of the categories as clearly inapplicable, and indicate the single category she considers most likely. She should note the

evidence that supports her estimate and the plausible conditions which, if inconsistent with her own implied model of project dynamics, would lead to a different conclusion.

Initially the standards committee should review all evaluations completed by members of EI. Even if at first the guidebook focuses only on three or four sectors, it is sure to miss important dilemmas facing evaluators in these areas and to provide guidance that is sometimes off the mark. Also the lines between serious and trivial violations of its principles and conventions may be hard to draw, and they will inevitably contain shades of gray. Normally no competent member of EI who makes a good faith effort to apply the EI approach should have reason to fear expulsion. The threat of expulsion is to ensure that EI evaluations are adequately designed to address the evaluation questions a project poses and to guard against major errors in impact estimates, particularly those that could represent bias. A member could only be expelled by a majority vote of the standards committee after she has had the chance to defend herself. The EI stamp, then, will be a signal that an evaluation has a credible design, that it is informed by relevant knowledge and experience from other projects, that it offers a defensible estimate of impacts and cost-effectiveness within an appropriate range, and that it gives a credible account of the causes of project impacts based on available information about the intervention's design and management. The guidebook's principles and conventions and guidance for their application will be improved over time.

Evaluation International's Contributions to Learning and to Accountability

Development professionals should be held accountable for learning from experience, and EI provides the analytic material for both the accountability and the learning. First, all EI

evaluations estimate the intervention's impacts and cost-effectiveness. Second, because impact estimates are based on consistent conventions they provide a basis for cross-project comparisons. Third, each evaluation explains how project results, whether favorable or unfavorable, were achieved. Hence EI evaluations will provide something the development community heretofore has lacked: a basis for identifying more cost-effective approaches to improving economic and social conditions in developing countries. Initially impact comparisons will be most reliable within sectors, but as conventions for estimating impacts are improved, cross-sector comparisons will also be supported. Also, fourth, EI evaluators will be familiar with the literature on the sector and with other evaluations of similar interventions. In due course their judgments and explanations will reflect the cumulative experience of the development community.

Most immediately, therefore, EI evaluations will be more analytically rigorous and will give better informed conclusions and recommendations. This will increase the contributions from evaluation to "local" learning, as evaluation findings inform ongoing program management. For donor agencies and the development community at large, EI will also support the identification of more cost-effective program designs and design features. With his famous 1980 paper, David Korten popularized the idea of a "learning process approach" for development agencies, and many international nonprofits subsequently came to describe themselves as learning organizations.² These agencies have largely depended, however, on their own in-house program review processes (often with external evaluations as material). EI will set the stage for a better informed learning process within individual agencies and across the development community.

² David C. Korten, 1980, "Community Organization and Rural Development: A Learning Process Approach," *Public Administration Review*.

Note that the analytic move from impacts to cost-effectiveness involves factoring in costs in a way that is consistent between evaluations. By committing to have EI conduct its evaluation, therefore, a development agency will give its staff an incentive to look for cost savings and for cases where high quality inputs, even if costly, can contribute to substantially greater impacts. As more evaluations are completed, there will be new opportunities to assess the merits of different management approaches and other design features. It will be possible to assess, for example, how more participatory approaches or approaches that make greater use of advanced technologies tend to affect a project's cost-effectiveness. Since the archive will be indexed, this will facilitate exploring such questions with statistical studies based on large numbers of projects as well as with more nuanced and qualitative multi-case comparisons.

Initially EI evaluations should focus on relatively discrete interventions, such as most development projects, where the intervention is the largest single factor influencing changes in the relevant conditions of the beneficiary population. All development interventions should aim to maximize cost-effectiveness given the appropriate conception of impacts and frame of reference (e.g. including qualitative impacts, and a wider frame for highly experimental projects). However when the variation in the relevant conditions caused by other factors exceeds the variations due to the project, such as with most structural adjustment programs and many sector-wide programs, and when causal connections between the intervention and changes in conditions are particularly hazy, it is harder to make reliable impact estimates. Evaluators must of necessity focus on intermediate variables of questionable significance, and expert judgment (which may not be reliable) often has to play a larger role. We can expect that through the operation of EI, the boundary of program types for which impacts can reliably be estimated will gradually

be extended. There are some kinds of programs, however, such as short term executive education programs, which will continue to resist evaluation on the basis of their impacts.

There are three main ways in which establishing EI is likely to enhance accountability. First, development planners can be held accountable for applying the lessons of experience in their program and project designs. Although the development community has been carrying out projects for more than 60 years, many planners are familiar with only a tiny slice of this experience and many projects are launched with weak designs. Besides estimating cost-effectiveness, EI evaluations will routinely explain how impacts unfold from project design and implementation. When cost-effectiveness is low, the evaluation will explain how far this is attributable to the project design. Heretofore donor agencies have tended to reward efficiency in committing funds or “moving money,” (Tendler, 1975; World Bank, 1992; Clements, 1993, p. 1634, pp. 1639-1642; Nelson, 1995; Clements, 1996; Horta, 2006, p. 9, Svensson 2006, p. 121), but once information on the contribution of project designs to impacts is available, it would be surprising if donor agencies failed to use it in personnel evaluations.

Second, project managers can be held accountable for modifying the project design in a timely manner in light of its unfolding experience. My own experience reviewing projects and their evaluations suggests that significant gains could be realized through more timely mid-course corrections. Probably the majority of project designs turn out to require substantial modifications, and evidence of mistaken assumptions or unrecognized constraints is often available quite early in a project’s lifetime (see e.g., Clements, 1999, p. 1376). Significant design changes, however, often involve changes in a project’s management structure, with implications for budget allocations and for personnel. Approvals may be needed from project principals in donor agencies and in the host country government.

While the learning that EI stimulates will help managers to identify when design changes are needed, anticipating an EI evaluation at project completion will also alter managers’ incentives. It will be apparent, at the evaluation, when evidence of a significant flaw in the project design became available. EI will provide a basis for holding managers accountable for taking such information into account.

Third, due to their greater rigor and independence, EI evaluations will uncover more of those (relatively few) instances of significant corruption. Large scale corruption inevitably has programmatic consequences, and evaluators accustomed to a cost-benefit framework will be sensitive to inconsistencies in project documents and curious when costs appear disproportionately high. Estimating project impacts requires the evaluator to look rigorously forward to likely project results, and this leads to a similarly rigorous investigation backwards into their causes. Many of today’s evaluations are relatively superficial and pay little attention to costs, and many evaluators are unfamiliar with procedures for initiating criminal investigations. EI evaluators, however, will be prepared to inform the appropriate authorities when they encounter evidence of fraud, so when EI is first established the number of prosecutions is likely to rise. After some time this will raise the perceived costs of engaging in corruption and its incidence can be expected to decline.

Establishing Evaluation International

Three pieces must be in place for EI to be established: evaluators to found it, donors and possibly implementing agencies to use its services, and resources to cover its start-up costs. A group presumably made up largely of evaluators will need to write the guidebook, develop and organize the examination, establish the standards committee, initiate the archive, and work out EI’s operating procedures. Before

the evaluators can proceed very far it seems that several donor agencies will need to commit to retain them for sections of their portfolios. The founding donors and evaluators together should select the sectors for the guidebook's initial focus and arrange for evaluations of a defined part of the donors' operations in these sectors to be carried out by EI members. As suggested above, the sectors might include agricultural extension, primary health care, water and microfinance. To protect EI's independence it would be better if its startup costs were not be covered by the founding donors. Perhaps private foundations or governments through the United Nations could cover these costs. EI can be expected to enhance the development community's professionalism and cost-effectiveness and hence speed achievement of the Millennium Development Goals and other reductions in world poverty, so it would be a reasonable investment.

Ideally EI's members will be drawn from across the countries that are donors and recipients of aid. The first priority, however, is to establish the competence of the association's membership. Since the depth and competence of the evaluation profession varies significantly across countries EI's initial membership may be skewed in its national representation, but every effort should be made to include wide participation from developing countries. In the long run sources of members will depend on how training and testing are carried out; these should be organized to secure balanced representation.

Assuming EI's appropriate establishment and its members' competence, we can expect EI evaluations on average to be technically superior to those of any donor agency today. What a donor gives up in control by joining EI it will gain in credibility and effectiveness. The launch of EI will be more robust if its initial members include a distribution of bilateral and multilateral donor agencies. While EI's donor members should be represented on its governing board, to defend EI's independence,

and in order that EI can promote the vitality and integrity of the evaluation profession, a majority of its board should be evaluators.

One task for EI's donor members will be to establish a scheme for valuing impacts that they can collectively support. Recall that each EI evaluation estimates project impacts and cost-effectiveness. Impacts are typically measured in units based on the kind of impact, but cost-effectiveness is measured in monetary terms. In order to compare projects with different kinds of impacts in terms of their cost-effectiveness, it is necessary to establish the value of each kind of impact on a common scale. This should not be thought of in terms of the inherent value of the impact (e.g. the value of a life saved, the value of a person having adequate access to clean water), but rather as how much donors are willing to pay to achieve a particular impact (Clements, 1995). Traditional economic cost-benefit analyses have assessed impacts in terms of their economic values at market prices or based on some notion of beneficiaries' willingness to pay, but these scales tend to neglect part of the value of non-economic impacts and they lead to implicit biases in favor of beneficiaries who are less poor. For EI evaluators to make consistent estimates of the cost-effectiveness of different kinds of projects they need to know how much value to assign to a given quantity of each kind of impact. Since it is donors who allocate funds to projects, it falls to them to specify how much the impacts are worth to them. It would enhance the evaluation community's unity and efficiency if the donors could agree on a single scale of values.

EI will clearly need quite a few seasoned evaluators among its founding members. As noted above, first they and the donor members need to select the initial sectors on which EI will focus. To write the guidebook they should consult texts on cost-benefit analysis and review a number of evaluation reports from the selected sectors. The guidelines, principles and conventions for estimating impacts should be worked out both deductively and inductively.

The deductive approach starts from the relevant impacts and one works out how they can best be estimated. The inductive approach starts from examples of stronger and weaker evaluations and one works out the principles (etc.) that are implicit in the former and the mistakes to be avoided from the latter. The two approaches should be combined especially for working out conventions for estimating impacts.

The founding evaluators will also need to establish reading lists for each sector. They should consult with sector specialists, but several of the evaluators should also read or re-read the texts. They should aim for a minimal list that provides an adequate basis for mastery of the dynamics of the sector. The examination to establish qualifications for entry to the association might include optional sections associated with each sector. In order for the association to build its character and organizational culture, the founding evaluators will need to establish a shared body of knowledge and sufficient agreement in judgments. Much of the latter may be worked out in formulating the entrance examination and particularly the principles and guidelines to govern operations of the standards committee.

The standards committee anchors the association's principles and hence its independence. The founders should determine the kinds of technical knowledge to be found among the committee's members, and selection procedures should ensure that these areas are covered. It is essential, however, that the committee should always possess the confidence of a majority of members. Hence its chairperson and two vice-chairs should be elected by the evaluator members, perhaps for terms of three years, and with some recall procedure.

The founding members can also build their foundation of common principles by commenting on evaluation reports to be included in the archive. The indexing system for the archive is mainly a technical matter,

although members should be encouraged to recommend terms for the index. The archive should be conceived, however, as representing a body of the evaluator members' professional judgments. As part of establishing the association the founding members should discuss the merits of a selection of evaluations, perhaps in an online forum. After completing the discussion the selected evaluations and their final commentary could launch the archive.

Conclusion

It should not be imagined that EI is opposed to methodological diversity in evaluation or that all EI evaluators should be alike. The EI approach to estimating impacts and cost-effectiveness should serve as grounds for a core analytic consistency among EI evaluations, but once this is established it is appropriate and desirable for diversity in methods and in areas of excellence to flourish among the association's members. This paper aims to suggest a few aspects of how EI might be established in order to clarify the basic proposal. In the actual event other issues are likely to be found more important and unanticipated constraints may undermine the relevance of some of these ideas. The basic proposition is that consistency in technically sound estimates of impacts and cost-effectiveness should provide a new foundation for professionalism in development management. The details of how this is achieved are less important.

Even though EI evaluations have characteristics of public goods, once EI is established its recurring costs should be borne by the users of its services. Evaluator members might pay membership dues which they recover in their evaluation fees and/or donors might contract for sets of evaluations with the association. Also donor members might pay subscriptions to sustain the archive.

The need for EI derives largely from the structural situation in which donor and implementing agencies, governments, and

powerful interest groups inevitably have the main influences on the organization of aid. The taxpayers and contributors who provide the resources for aid and aid's intended beneficiaries in developing countries, by contrast, have little power to defend and promote their interests. In this context it is appropriate to view evaluators as defending the interests of the poor by developing and adhering to high technical standards in estimating program and project impacts and cost-effectiveness. Strong evaluation in some measure "corrects for" the disempowered status of aid's intended beneficiaries.

Of course evaluation by itself cannot complete the loop of learning and accountability. It will be up to other members of the development community, especially officials of donor and implementing agencies, to use evaluation findings in their personnel management. Once managers are rewarded for effective management and penalized for failing to maintain professional standards, however, this will alter the incentive environment for development professionals generally. Given stronger incentives and better materials for learning, development professionals will tend to make more cost-effective choices, and development resources will be better used to reduce world poverty.

References

- Bollen, K., Paxton, P. & Morishima, R. (2005). Assessing international evaluations: An example from USAID's democracy and governance program. *American Journal of Evaluation*, 26(2), 189-203.
- Chianca, T. (2007). *International aid evaluation: An Analysis and policy proposals*. Unpublished doctoral dissertation, Western Michigan University, Kalamazoo.
- Clements, P. (1993). An approach to poverty alleviation for large international development agencies. *World Development*, 21(10), 1633-1646.
- Clements, P. (1996). *Development as if impact mattered: A comparative organizational analysis of USAID, the World Bank and CARE based on case studies of projects in Africa*. Unpublished doctoral dissertation, Princeton University, New Jersey.
- Clements, P. (1999). Informational standards in development agency management. *World Development*, 27(8), 1359-1381.
- Horta, K. (2006). The World Bank's decade for Africa: A new dawn for development aid? *Yale Journal of International Affairs*, Winter/Spring, 4-23.
- Korten, D. C. (1980) Community organization and rural development: A learning process approach. *Public Administration Review*, 40(5), 480-511.
- Kusek, J. Z. & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system*. Washington, DC: The World Bank.
- Nelson, P. J. (1995). *The World Bank and non-governmental organizations: the limits of apolitical development*. New York, NY: St. Martin's Press.
- Scriven, M. (2007). *Key evaluation checklist*. Retrieved February 13, 2008 from http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf
- Svensson, J. (2006). The institutional economics of foreign aid. *Swedish Economic Policy Review*, 13, 115-137.
- Tendler, J. (1975). *Inside foreign aid*. Baltimore, MD: Johns Hopkins University Press.
- World Bank. (1992). *Effectiveness implementation: Key to development impact*. The Wapenhans Report. Washington, DC: The World Bank.
- World Bank. (2007). *World Bank assistance to agriculture in Sub-Saharan Africa: An IEG review*. Independent Evaluation Group. Washington, DC: The World Bank.