Monitoring and Evaluation for Cost-Effectiveness in Development Management

By Paul Clements

Recent efforts to focus aid on countries with favorable economic policies have been motivated in part by the fact that cross-country econometric studies have failed to find a causal relationship between development assistance and economic growth. This can lead to a neglect of many of the neediest, however, as many of the poorest countries also have poor policies. An alternative account of aid’s limited impacts can be found in weaknesses in monitoring and evaluation systems. The on average positive findings from project evaluations are hard to reconcile with the macroeconomic evidence, and the structural conditions of development assistance make evaluation vulnerable to positive bias. While much of the evaluation literature focuses on attaining statistically valid impact estimates or on participatory approaches that empower beneficiary populations, this essay presents an approach to monitoring and evaluation that aims to strengthen judgments of cost-effectiveness. The proposed approach involves evaluators achieving independence from project management in a manner similar to that of accountants in the private sector.

I. INTRODUCTION

Recent decades’ efforts to reform development assistance have been oriented largely (1) to improving the economic incentive environments in developing countries in order to improve resource allocations and to stimulate entrepreneurship or (2) to increasing popular participation in project management or, more broadly, in governance. As important as these reforms have been (and remain), they have tended to locate “the problem” outside the traditional provenance of development assistance. Granted participatory reforms have sought to change the way development assistance is carried out. These reforms have often been conceived, however, as additional to or as displacing mainstream practice. Efforts to improve the incentive environment, such as through structural adjustment, have sometimes been understood as an alternative to supporting government in carrying out development programs. The approach I should like to propose, by contrast, aims to reform traditional development practice, as it were, from the inside out.

The traditional core task of development assistance has been, and largely remains, the implementation of programs and projects that aim to improve economic and social conditions. It has been thought that prior advances of advanced industrial countries indicate a potential for improvements in countries that remain poor. Development management, in this view, involves identifying a promising investment, securing funding, forming an implementation team, taking steps to give material form to the project idea, establishing relationships with beneficiaries and other stakeholders that elicit
participation and support, and modifying the plan as necessary in the face of unanticipated events so that the project becomes institutionalized, yielding a flow of benefits more than commensurate with costs, within the social fabric. This unexceptional sequence, it turns out, requires a particular form of consciousness, a way of thinking about the project and its likely consequences, that can be specified in its broad outlines. It involves conceiving of the project and its evolving relation with its environment as a unified, organic whole, with a perspective grounded in an understanding of likely project impacts. Experience suggests that this form of consciousness, although far from universal among development practitioners, is also not uncommon. More importantly, it can be learned and trained. It can provide the orientation for an organizational culture (or organizational culture can be inhospitable or hostile to it). I shall argue that monitoring and evaluation systems provide tools that can establish and maintain this kind of consciousness. The practice of monitoring and evaluation in the development field today, however, falls far short of this ideal.

To see the problem from another perspective, we can say that the structural conditions of international development assistance place an unusually heavy burden on monitoring and evaluation systems. Consider how the pressures and demands on resource allocation decisions (human and financial) in international development organizations compare with those facing managers in other kinds of organizations. A firm in the private sector has an incentive to serve customers because they pay for the product. The desire to maximize profits and the need to compete with other firms present consistent management imperatives. A government agency, at least in a democracy, typically faces a less direct but nevertheless potentially powerful form of accountability to clients by way of the ballot box. Democracy hopes to produce elected representatives who are good stewards of the public interest, and citizens dissatisfied with the performance of a government agency can seek reforms through the democratic process. The structural accountability of managers in international development organizations to their clients, by contrast, is extraordinarily thin. On the contrary, development agencies are likely to be influenced by the democratic process to support the economic and political interests of the donor countries. We can view monitoring and evaluation systems as filling the resulting gap: as an artifice to establish a form of accountability to beneficiary interests that structural conditions cannot provide.

The essay is organized as follows. Section II addresses the scope of the problem, reviewing evidence that aid’s contributions to economic growth have been limited and discussing standards of monitoring and evaluation (M&E) in the development field. Section III presents two prominent approaches to M&E in the development literature, approaches we will call “M&E for empowerment” and “M&E for truth,” and introduces M&E for cost-effectiveness as a (friendly) alternative. Section IV gives more details on the steps involved in implementing M&E for cost-effectiveness, and Section V explains why an evaluation association should be established to address the structural problem discussed above. Section VI speculates on results we might expect in the long term from the reforms the essay proposes.
II. ECONOMIC POLICY, PROJECT EVALUATION, AND THE IMPACTS OF AID

Lingering doubts about the effectiveness of development assistance have been exacerbated by disappointing evidence from cross-country econometric studies. Studies by Mosley, et al. (1987) and by Burnside and Dollar (2000) have failed to find statistical evidence that foreign aid contributes to economic growth in the average recipient country. Boone (1996) finds no significant contribution from aid either to economic investments (except in countries that receive more than 15% of GNP in aid) or to basic measures of human development such as infant mortality and primary schooling rates. Burnside and Dollar (2000) look at the relationship between the total grant component of foreign aid (bilateral and multilateral, excluding the loan component) and economic growth in 56 low and middle income countries from 1970 to 1993. After taking account of such factors as assassinations (a measure of political instability), institutional quality, the size of the budget surplus, inflation, openness to foreign trade, and whether a country is in Sub-Saharan Africa (indicating lower growth), Burnside and Dollar find no statistically significant relationship between the aid a country receives as a percentage of its GDP and its economic growth rate. This does not rule out the possibility that aid contributes to growth in some cases, or that aid has other positive impacts such as on health, education, or political stability.

Burnside and Dollar also find, however, that aid does contribute to economic growth in countries with “good” economic policies. They construct a policy index based on the budget surplus, the inflation rate (with a negative coefficient), and openness to trade, and find that when this index is sufficiently high there is a statistically significant, positive contribution from aid to growth. On this basis they argue that aid allocations should depend on economic policies, so governments that maintain low budget deficits (or some surplus) and a low rate of inflation, and that limit their controls on foreign trade, would get more aid. This would increase aid’s contribution to economic growth in recipient countries and give countries that are negatively affected an incentive to adopt good economic policies.

It is worrying that of the 56 countries in their study, only six had governments that maintained good economic policies (Botswana, Indonesia, Korea, Malaysia, Thailand and Turkey; Table A2, p. 866). By the final period of the study, 1990-1993, 15 of the study’s 40 poor countries had managed to establish good economic policies, but this was not enough to cause aid to show a positive impact on growth overall. Burnside and Dollar’s findings indicate that foreign aid contributed little to economic growth in Latin America or Africa (except in Botswana) during the 1970s and the 1980s.

It is somewhat surprising, in this context, that most bilateral and multilateral development agencies found their own projects in these periods to be overwhelmingly successful. As a long-time observer of foreign assistance, I have never seen a development agency report that finds the agency’s projects to have been generally unsuccessful. The most comprehensive study of aid evaluations from this period, Does Aid Work? Report to an Intergovernmental Task Force by Cassen and associates (1986; 2nd edition 1994), makes a typical finding.
In the broadest sense, this report finds that most aid does indeed ‘work’. It succeeds in achieving its development objectives (where those are primary), contributing positively to the recipient countries’ economic performance, and not substituting for activities which would have occurred anyway (1994, p. 7).

Cassen cites a World Bank review of over 1000 projects completed between 1975 and 1983 that found 79% with economic rates of return of 10% or greater, and around 90 percent of total investments

appeared to have achieved their major objectives, or were on the way to doing so. … The common conclusion [across development agencies] is that some 65-75 percent of projects are found to be satisfactory or highly satisfactory, and most of the remainder problematic but not irreparably so, with a small percentage (in single figures) completely written off (1994, p. 90).

The only way this would square with Burnside and Dollar’s findings would be if most investments did not aim to contribute to economic growth and/or if aid displaced government investments that would have been just as profitable, leading instead to unproductive government spending. But despite the poverty alleviation rhetoric of the 1970s, economic growth remained the dominant aid objective in this period. Most donor-funded development projects are so different from what governments would have done without the aid that it is hard to see how aid could have been giving governments the opportunity to switch spending from profitable investments to unproductive consumption.

Cassen notes that most donor agencies in the 1970s and ‘80s did not require estimates of economic rates of return for most of their projects (1994, p. 90). The World Bank imposed the discipline of requiring rate of return calculations because all its projects are funded by loans (at more or less concessional rates compared to what governments would face on the world market). A project would represent a net drain on the economy if it failed to achieve an economic rate of return greater than the value of loan repayments. There is a basic incentive problem, however, in the way economic rates of return are re-estimated at the conclusion of World Bank projects. It is officials who have been responsible for the project who are responsible for having rates of return re-estimated. On one hand this is an advantage. They know the project and are likely to be main consumers of the analysis in the Project Completion Report where economic returns are re-estimated. On the other hand, however, it is typically in the interest of the donor agency, of project management, and of the borrower government for their completed projects to be seen as successful. Even donor agency officials who have no connection with the project have an organizational incentive to report positive results. Most projects have only begun to recoup their investments at the time implementation is completed, so there is usually much ambiguity in rate of return estimates at this time. Instead of other donor projects being so much worse than World Bank projects in the 1970s and 1980s that they dragged down aid’s average economic impacts, it is plausible that there may have been positive bias in re-estimated returns for World Bank projects.
Evaluators of donor-funded development projects generally tend to be in a delicate position vis-à-vis the donor and project managers. Many evaluators are independent consultants, while some work for consulting firms that hold evaluation contracts and some are regular employees of a donor agency. In all cases their future employment prospects tend to depend to some extent on the donor. Some evaluations are carried out by teams consisting only of donor agency, government, and/or project employees. Such evaluations may be presumed to be vulnerable to (possibly unintentional) positive bias. Probably the majority of evaluations are led by someone who is not a regular employee of an agency responsible for the project, so evaluations tend at least to provide a fresh overall perspective. When evaluators suspect that their access to future contracts may depend on their presenting the present project in a positive light, however, their independence is likely to be compromised to some extent.

Several factors contribute to the incentive for and vulnerability to bias in program evaluations. Most obvious is the desire for one’s work to be seen as successful, and the possibility that one’s future access to resources may be diminished if one’s work is seen as unsuccessful (Cracknell, 2000, p. 184). Another source is disciplinary fragmentation, or the multiplicity of disciplinary perspectives among development professionals. Specialists in agriculture, engineering, health, economics, anthropology, and other areas bring important training and specialized knowledge to development tasks, but to the extent that they each understand development problems in the terms of their own academic disciplines, their perspectives on a particular program are likely to conflict. One often finds “camps” in a donor agency, with the agricultural extension people, the family planning specialists, the economists and so on defending their own turfs vis-à-vis one another (Cracknell, 2000, p. 184). Anecdotal evidence suggests that in some cases their defense of their own programs leads them to ignore or to hide unfavorable findings.

While the World Bank has estimated economic rates of return for most of its projects, most donor agencies have planned and evaluated projects in terms of (often hierarchical) lists of objectives. Economic analysis may be of limited value when a project’s goals are not in the first instance economic. Economic analysis does, however, have attractive properties for evaluation purposes. An estimated economic rate of return provides a summary statement of all a project’s economic costs and benefits, and there are fairly clear rules for estimating benefits. A project planned on the basis of a list of objectives may provide a similarly determinate basis for evaluation if objectives are quantified and the relative weights of different objectives are clear, but most such projects do not. Commonly there are half a dozen objectives, including a few (e.g. “increase incomes,” “strengthen community organizations”) that require significant judgment calls from evaluators. Since the overwhelming majority of projects make some progress on most if not all objectives, it is usually not too hard to present a project as a qualified success.

There are few independent analyses of evaluations of donor-funded projects in the published literature. Carlsson, et al., in The Political Economy of Evaluation: International Aid Agencies and the Effectiveness of Aid (1994), review evaluations of development projects in Nicaragua (gold mines), Kenya (soil conservation), and
Madagascar (forestry management), and of commodity import support programs. They find little rigorous economic analysis to support the economic rates of return estimated for these projects.

Most appraisal and evaluation manuals convey an image of a structured and rational decision-making process, but it seems as if the reality is something else. It rarely happens that a project idea is subjected to any economic analysis of substance, which then forms the basis of the decision making. … Even an appraisal system as rigorous as the World Bank’s is in practice continuously being manipulated, because it is subordinated to the individual interests of POs [project officers] (getting projects to the Board) as well as the organization’s own objectives (meeting the disbursement targets). … Individuals are rational in the sense that they defend their, or their group’s interests (Carlsson, et al., 1994, p. 180).

Clements (1999) reviews four US Agency for International Development and four World Bank projects in Uganda, Kenya and Malawi. Two of these projects had economic rates of return re-estimated when the project was completed, and both involved glaring methodological errors leading to much higher reported returns than valid analysis would allow.

Out of more than 10 applications of cost-benefit analysis found in project documents (often for individual project components), at the time of the study only that for [one project component in] road maintenance was clearly sustained by outcome information to give a valid conclusion about project results. Final reports were not yet available for five projects … but none of these was collecting a base of information that would support coherent overall impact analysis. Where there was no determinate evaluation framework such as cost-benefit analysis provides, positive reports were often achieved by identifying positive outcomes and ignoring or underreporting negative outcomes (Clements, 1999, p. 1377).

These studies suggest that rigorous analysis of project impacts is uncommon and positive bias probably the norm.

The repeated failure to find statistical evidence that aid contributes to economic growth is inconsistent with the positive summaries of donor agency project portfolios based on project evaluations. It appears that the statistical evidence is closer to the truth than the evaluation summaries. There are significant incentives and opportunities for positive bias in project evaluations, and the limited available evidence on assessments of project impacts (economic or otherwise) confirms the presence of positive bias. Monitoring and evaluation systems are not providing the kind of accountability that donor agencies need to orient their management systems to serving beneficiary interests. Yet the accountability function is only one function of evaluation. At least as important is the role of evaluation in organizational learning. As things stand donor agencies simply do not know which of their operations are cost-effective.
III. MONITORING AND EVALUATION IN DEVELOPMENT MANAGEMENT: THREE APPROACHES

Two different kinds of reforms are needed for monitoring and evaluation systems to enhance the cost-effectiveness of development assistance. Structural reforms that can address the problem of positive bias are discussed in Section V. Here we discuss an approach to monitoring and evaluation designed to support stakeholders for donor-funded projects in judgments of cost-effectiveness. In “M&E for cost-effectiveness” it is the role of evaluation to reconstruct and articulate the evolving hypotheses implicit in a project design in terms that universalize the project’s unique problems and solutions. Given that development management involves a constant competition for resources, and given the structural condition discussed above in which beneficiary interests are only weakly represented, it is the role of evaluation to make the strongest possible judgments about a project’s cost-effectiveness in terms that facilitate comparisons among projects. These two features – reconstructing the project hypothesis and making strong judgments of cost-effectiveness – together constitute a distinctive approach to monitoring and evaluation that supports the synthetic judgments on which effective development management depends.

To explain M&E for cost-effectiveness it is useful to compare it with other evaluation approaches. The strongest challenge to standard approaches to aid evaluation in the last two decades has involved the elaboration and application of participatory approaches (see e.g. Cracknell, 2000). These have aimed to involve beneficiary populations in project management, to assist them in taking responsibility for improving their own conditions and to incorporate them in more democratic processes of development decision making. Authors such as Korten (1980) and Chambers (1994a, 1994b, 1994c), whom Bond and Hulme describe as “purists” (1999, p. 1340), have sought to reorient the development enterprise to support the goal of empowerment. They have promoted an approach I call “M&E for empowerment” because it emphasizes learning at the local level, seeking to empower project beneficiaries by involving them in the evaluation process. While M&E for cost-effectiveness appreciates empowerment as an important development goal, it identifies the locus for the learning that evaluation should support among those who are responsible for resource allocation decisions. Donor agency officials are the primary audience for aid evaluation because they exercise primary control over these resources. It turns out, however, that the form of evaluation that can best inform these officials will also best inform officials of developing country governments, project managers, and the overall development community, as well as, with some additional synthesis, the legislatures that appropriate aid budgets. Evaluation and empowerment goals overlap in their management implications, and empowerment was certainly neglected by the development community prior to the mid-1970s. Monitoring and evaluation for empowerment should be viewed as a subsidiary branch of aid evaluation and as a management strategy for empowerment.

It is possible that a great practical barrier to useful evaluation arises from some of those most knowledgeable of and committed to evaluation as a science. It has been common practice to begin discussions of aid evaluation methodology with the
experimental method of the natural sciences (e.g. Casley and Lury, 1982; Baker, 2000), and to present the various evaluation methods as, in effect, more or less imperfect approximations to randomized and controlled double-blind experiments. This approach often uses household surveys that measure conditions that a project seeks to influence, so that through appropriate comparisons changes attributable to the intervention can be identified in a statistically rigorous manner. I call it “M&E for truth” because it emphasizes making statistically defensible measurements of project impacts. This approach is right to insist that projects should be assessed primarily on the basis of their impacts, and that impacts should be understood as changes in the conditions of the population compared to what would be expected in the project’s absence (in evaluation jargon, as compared to the counterfactual). It is arguable, however, that in its orientation to statistical rigor it has established a “gold standard” that many evaluators are all too quick to disavow. Only a very small proportion of project evaluations present statistically rigorous impact estimates, and evaluations that do not often use the demanding requirements of statistical rigor as an excuse not to address the question of impacts at all.

Monitoring and evaluation for cost-effectiveness differs from M&E for truth in its view of the role of evaluation and in its methodology. Managers allocating resources for development assistance, which involves investment choices in many sectors and in dozens of countries, must come to terms with staggering complicated and difficult problems. To maintain a working understanding of the strategic outlook even for a few projects in one country is an enormous challenge. To support the judgments these managers have to make, M&E for cost-effectiveness first has evaluators not only estimate impacts but also make judgments of cost-effectiveness. Second, this approach aims as much to achieve a clear and succinct account of a project’s strategic developments, or cause and effect relations, as to enumerate its impacts. A great challenge to making judgments of cost-effectiveness lies in the weighting of different kinds of impacts. Evaluation of cost-effectiveness has generally been understood to involve assessing whether a particular, usually non-economic objective has been achieved at the least cost or at a reasonable cost (e.g. Gittinger, 1982, pp. 280-284). Problems in development management, however, routinely require choices among unlike goods. The distinctive feature of judgments of cost-effectiveness (as understood here) is that they are highly synthetic. They involve summarizing over a range of outcomes and impacts in terms of a standard imposed by the cost constraint. If the problem of weighting can be adequately addressed (see Section IV), the justification and defense of judgments of cost-effectiveness can be expected to contribute substantially to bureaucratic learning in development management.

To know that project X has achieved impacts Y, even if Y is a complete and accurate summary of impacts, is not necessarily very helpful to subsequent resource allocation decisions. One also needs a textured understanding of the overall features of the project design, of relevant features of the beneficiary population, and of management challenges that have been faced and their disposition (as well as information on costs). The lessons from a project are in its dynamics, not in its results. Information on impacts is needed to mark the nature of the lessons to be learned (e.g. of success or failure) and to
ground and buttress the analysis of management strategy. Therefore M&E for cost-effectiveness places as much importance on analyzing a project’s unfolding hypothesis as on analyzing its impacts and cost-effectiveness. It aims to explain the unique features of the project’s experience as a development intervention in terms that support the application of lessons learned to other possible projects. A final evaluation, under this approach, should include a reconstruction, beginning with the project design and relevant features of the beneficiary population and of the project environment, discussing the main components and features of project implementation, any changes in assumptions or design, problems that have been faced, and so on.

Besides taking this holistic approach to evaluation, M&E for cost-effectiveness also adopts an integrative methodology. This methodology starts from the kinds of judgments managers are required to make. It views the particularly strong forms of evidence that statistical methods can generate as planks to be used in constructing judgments of impacts. Household (and other) surveys, however, although they should not be disparaged, are not the only defensible basis for impact estimates. A project design itself presents an analytic framework from which the question of impacts can be approached, and project records and opinions from project managers provide a starting point for developing a composite picture of changes due to a project. The propositions of a project design can be probed and its picture of results elaborated through surveys, interviews, direct measurements, case studies, comparisons with other projects, and other investigative techniques to test how far the changes intended in the project design have taken place. Under M&E for cost-effectiveness, impact estimates are based on many forms of quantitative and qualitative information. Scenarios can be developed, for example, of “high,” “medium,” and “low” impacts, and it may be possible to determine with no ambiguity that actual results are incompatible with all but one of the scenarios. This approach is better described as logical and integrative than as scientific, because we judge the strength of its conclusions not by standards of statistical validity (although statistical data that it employs should be judged by these standards) but with our normal (trained) human reason. The test of an evaluation here lies in how cohesively it analyzes the intervention as an intervention of its type, how it marshals evidence, how it assesses problems presented in the implementation of the project, how coherently the judgment of cost-effectiveness summarizes the project experience, and how consistent this judgment is with appropriate standards of judgment, possibly represented by how similar judgments have been made for similar projects.

Monitoring and evaluation for empowerment emphasizes participatory methods whereby beneficiary populations can reach, or contribute to, evaluative judgments. Monitoring and evaluation for truth emphasizes statistical methods whereby conclusions about project results can be defended when they have been reached by steps consistent with statistical rules. Monitoring and evaluation for cost-effectiveness emphasizes assessments of project designs and unfolding strategies grounded in judgments of cost-effectiveness. It is no mistake that M&E for cost-effectiveness places so much emphasis on broadly integrative judgments, as there is much scope for foreign assistance to be improved if these can be strengthened. Various ways in which this can be accomplished are discussed in the remaining three sections.
IV. MONITORING AND EVALUATION FOR COST-EFFECTIVENESS

Over the last ten or fifteen years there has been a trend among United Nations and bilateral development agencies and the private contractors and non-governmental organizations (NGOs) they fund to increase the extent to which program assessments are based on outcomes rather than on outputs. Agencies that require estimates of economic returns for most of their projects are not part of this trend, as such estimates require a summary at least of economic outcomes. Indeed it is ironic that the World Bank and other multilateral development banks seem to have de-emphasized the economic analysis of projects during this period. Monitoring and evaluation for cost-effectiveness can be seen as continuing and intensifying the emphasis on outcome-based assessment. It adds a comparative dimension, as judgments of cost-effectiveness imply a comparative perspective and can be supported by actual or implicit comparisons with similar projects. Consistently with the trend, it aims to strengthen the focus on impacts in project plans, and it favors tightly constructed, outcome-oriented monitoring systems. Monitoring and evaluation for cost-effectiveness has the additional ambition, however, to place economic cost-benefit analysis and other approaches to outcome-based planning and evaluation (such as the logical framework) under a unified framework.

The motivating idea is that project managers and other management stakeholders are in a better position to make resource allocation decisions when they have a model of the project and its likely impacts in their minds. Also, they will be better able to orient management decisions to cost-effectiveness when their mental map of the project is oriented in this way. A credible estimate of likely project impacts requires an analysis of the project design, the changes it expects to bring about in beneficiary behavior, the relevant features of the environment, and the conditions that must be maintained for the anticipated improvements in beneficiary conditions to be realized. Estimating impacts requires individuals to think through project dynamics in a particularly thorough way, and it allows management stakeholders collectively to build a strong foundation for resolving conflicts and justifying choices. Management stakeholders are collectively responsible for the project’s long term results, and to estimate impacts is to engage in the central form of deliberation that is essential to discharging this responsibility.

It is not to be expected that impact estimates will generally be accurate. We should not think of managers’ organic conception of the project even ideally as the true conception, but, ideally, as a very sound conception (for management purposes) in light of available information and experience. One way to see the significance of M&E for cost-effectiveness is to imagine what happens when management stakeholders discover that part of their model is false. If they have an organic analysis of the project in their minds they will be drawn to see which features of the project design are called into question. They will think through changes in the project design that could accommodate the newly perceived condition. If their analysis is oriented to cost-effectiveness, then if the situation is so drastic that plans for the unspent portion of the project budget are no longer likely to be cost-effective, they will consider halting the project. Perhaps they will revise the entire strategy. An analysis grounded in the idea of cost-effectiveness provides
a background against which we can rightly say that new information “calls for” specific changes in strategy.

The most important steps are (1) for project plans to be constructed around an estimate of project impacts and cost-effectiveness, and (2) for an evaluation to be carried out at the conclusion of each project that updates these estimates. Also a monitoring system that tracks parameters in these estimates should be implemented during the life of the project. Every project plan begins with some idea about how actions X can lead to changes Y. To build a plan around impact estimates is to project forward, as the details of the plan are worked out, the changes in beneficiary conditions (as relevant to development) that can be anticipated due to project activities.

An estimate of project impacts is made through some combination of internal logical construction and application of lessons from other contexts. It normally takes the form of a list of conditions, and for each condition a quantitative and/or qualitative estimate of the extent of change. Although there can be much uncertainty in impact estimates, particularly regarding the duration for which benefits should be attributed to a project, the more controversial matter involves the judgment of cost-effectiveness. The difficulty is that there is no consensus on how changes in different conditions can be weighted, or compared, and combined or summed together for a single judgment of cost-effectiveness. The answer I propose involves a mixture of legislative guidance, professional discretion, and a long-term collective consensus-building process.

The problem of weighting benefits is not new to the development community. In the 1970s the World Bank adopted the goal to reduce poverty, but in standard economic analysis whether benefits accrue to rich or poor individuals is irrelevant. All that is counted is the sum of benefits. Therefore the Bank developed distributionally sensitive cost-benefit analysis, where economic benefits accruing to individuals with lower incomes count for more according to a distribution weighting factor (Squire and Van Der Tak, 1975). Although it was not widely used, in this system it was the government taking the development loan that would set the rate of distributional sensitivity (p. 63). Clements argues that when loans are highly subsidized, as most World Bank loans to low income countries are, it should be the governing body of the lending agency that makes this judgment, as this would support a consistent focus on poverty across countries (1995, p. 582). Also, if different evaluations apply greatly different weights to the same kinds of benefits it makes it harder to generalize across evaluations and to learn from their conclusions.

For these reasons it is advisable for governing bodies such as the World Bank’s board of governors, or, for bilateral agencies, the donor country legislature, to set guidelines for comparing, for example, income gains to poor and rich households, gains in life expectancy, and the attainment of literacy. If the legislature can identify the sums it approves spending for different kinds of impacts then comparisons can be made in monetary terms. For many projects, however, there are significant impacts that cannot be expressed in standard units. In order that these should not be excluded from the analysis
of cost-effectiveness there should be some scope for planners and evaluators to assign values to project impacts.

Under M&E for cost-effectiveness the initial estimate of cost-effectiveness is made when the project is planned. Planners are to specify not only the impacts they expect the project to yield, but also, given expected costs, how cost-effective they expect the project to be. We can use the rating system employed by economic analysis as a conceptual anchor for a scale of cost-effectiveness (see Figure 1). An economic cost-benefit analysis yields an economic rate of return (ERR) that expresses the value of a project’s economic benefits as a return on the investment. The World Bank has traditionally set a benchmark of 10%, expecting projects in the economic sectors to achieve ERRs at least this high. An ERR above 20% is very good, while returns below 5% are disappointing. A scale of cost-effectiveness based on this system is presented in Figure 1. Here a rating of 6, corresponding to an ERR above 30%, is excellent, and a rating of 1, corresponding to an ERR below 0%, is disastrous. Planners list type, quantity (and/or quality) and weight for the impacts they expect the project to yield, with weights either supplied by the legislature or drawn from their professional judgment. Taking costs into account, they locate the project on the scale of cost-effectiveness (from 1 to 6). Then they identify plausible or near-plausible variations in impacts which, given the same costs, would move the project up and down one unit on the scale (assuming it is not already a 6). In this way they establish an evaluative framework that applies directly to the project’s conditions.

![Figure 1: Scale of Cost-Effectiveness](image)

<table>
<thead>
<tr>
<th>Economic Rate of Return</th>
<th>Degree of Cost-Effectiveness</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% and above</td>
<td>6</td>
<td>Excellent</td>
</tr>
<tr>
<td>20% - 29.9%</td>
<td>5</td>
<td>Very good</td>
</tr>
<tr>
<td>10% - 19.9%</td>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>5% - 9.9%</td>
<td>3</td>
<td>Acceptable</td>
</tr>
<tr>
<td>0% - 4.9%</td>
<td>2</td>
<td>Disappointing</td>
</tr>
<tr>
<td>Below 0%</td>
<td>1</td>
<td>Disastrous</td>
</tr>
</tbody>
</table>

When the project is completed these estimates are updated based on actual results. If project managers have been maintaining their own ongoing analysis of project impacts and cost-effectiveness, then the evaluation at project completion should hold no surprises for them. It is important, however, for final evaluations to support learning not only by those involved with the project, but also by the funding agencies, the host country government, and the development community at large. The main requirement for this purpose is that estimates of impacts and of cost-effectiveness should be methodologically consistent across evaluations. Indeed, one of the main innovations proposed under M&E for cost-effectiveness is the establishment of an evaluation association that, among other
functions, could support its members in employing consistent methods and approaches in the conduct of their evaluations (see Section V). The evaluation provides a summary of the initial project design and an account of how the hypotheses implicit in this design unfolded over the life of the project. The overall justification for the project, however, and for development assistance overall, rests on project impacts. The lessons to be learned from the analysis of project dynamics depend on the assessment of impacts and of cost-effectiveness. Impact estimates provide the main anchor not only for project evaluations, but also, due to structurally weak accountability to beneficiaries, for development assistance as a professional venture. We can say that the lack of methodologically consistent impact estimates in evaluations heretofore has undermined the professionalism and integrity of the development assistance community.

The need for methodological consistency can be seen from experience with economic cost-benefit analysis. The estimation of an ERR when a project is completed typically depends largely on benefit streams that continue well into the future. Such long term projections, however, can be very sensitive to fairly minor changes in assumptions. For example, it may be that established organizational capabilities and institutional arrangements are not yet strong enough to ensure the maintenance of new infrastructure, but a follow-on project that aims to improve these capabilities and arrangements has been designed. If a final evaluation assumes that the follow-on project will succeed, it may give a much higher ERR than one based on the present situation. For estimates of cost-effectiveness to be useful they must be comparable across projects, and for this they must be consistent in the way they resolve this kind of question.

Even if methodological consistency can be achieved, given the current state of the evaluation art and the limited resources available for evaluation there will be many cases in which impact estimates and therefore judgments of cost-effectiveness cannot be very precise. In light of the structural burden on evaluation in development assistance, however, as well as the need, given the difficulty and complexity of development assistance, for learning between projects, it is important for evaluators to make the strongest judgment of cost-effectiveness that they consider justifiable. In cases where available information does not permit a confident judgment, they should still make their best guess. They should indicate their level of confidence in the guess (e.g. high, medium, low), and the range on the scale in which they are confident that the project’s likely degree of cost-effectiveness lies. For example, evaluators may conclude that they have medium confidence that Project X has a degree of cost-effectiveness of 3, but they are highly confident that it is either a 3 or a 4 (acceptable or good). In this case they should indicate the factors which, if known, would allow them to make a confident decision, and show how plausible values of these factors affect impact estimates and the judgment of cost-effectiveness.

It would serve the development enterprise in various ways if final evaluations like this were carried out routinely. It would provide a basis for holding project planners and managers accountable for making proper use of the resources under their control, making up, to some extent, for the structural constraints to beneficiaries holding them accountable. It would provide donor agencies and the development community with
ongoing feedback not only on which kinds of projects are more cost-effective, but also on strategic and environmental factors that contribute to cost-effectiveness. In this way it would support ongoing improvements in project portfolios. We could expect that project planners would read evaluations of recently completed projects similar to the one they have in mind, as this would aid in project design and in estimating likely cost-effectiveness. We could also expect that project managers would read them, and, before accepting responsibility for a particular project, determine to their own satisfaction that impact estimates in the project plan are plausible. It would support the entire development community in constructing and maintaining a background orientation to cost-effectiveness. This would enhance the rationality of development assistance as an enterprise in pursuit of development goals.

V. AN ASSOCIATION FOR INDEPENDENCE, CONSISTENCY, AND METHODOLOGICAL ADVANCES IN THE EVALUATION OF DEVELOPMENT ASSISTANCE

It is unlikely that M&E for cost-effectiveness can be established simply from a specification of its steps and an explanation of its advantages. Positive bias in contemporary practice stems not primarily from methodological shortcomings among evaluators, but from a conflict of interest inherent in the structural conditions of development assistance. Short term interests of management stakeholders in their individual and organizational reputations conflict with long term interests of the development community in accountability and in learning from experience. Comparing the situation of the intended beneficiaries of foreign assistance with that of the clients of most organizations in the public and private sectors (see Section I), the weakness of development “clients” is the other side of the coin to the vulnerability of development project evaluations to biases arising from management interests. The apparent prevalence of positive bias suggests that organizational cultures and procedures for evaluations have evolved to accommodate the interests of management stakeholders.

When evaluators are paid from funds originating from donor agencies, (an apparently permanent condition,) there tends to be an incentive for evaluators to present their findings in a positive light. One way for evaluators to achieve a significant degree of independence from the interests of management stakeholders, and to establish an independent frame of reference, is to establish an evaluation association with its own rules and standards. The structural problem facing evaluators in development assistance parallels the situation of accountants and auditors in publicly traded corporations. The interest of corporate managers in using corporate resources for private purposes conflicts with shareholders’ interest in managers promoting profits. Yet the accountants who keep the managers’ books and the auditors who check that the books are kept properly are employed by the managers. Accountants and auditors are protected from the private interests of management by the rules of their professions, as codified, for example, in Generally Accepted Accounting Principles (GAAP), and as interpreted by associations of their peers. These structural devices are not always effective, as recent ENRON and WorldCom scandals demonstrate, but such scandals are the more remarkable because the controls usually do work. A professional association of development project evaluators
would not eliminate incentives for positive bias, but it would provide a counterweight, while also promoting technical improvements in the way evaluations are conducted.

The independence that a professional association could provide would be grounded in the rules and professional standards that it maintains. Monitoring and evaluation for cost-effectiveness could hardly be sustained without a professional association (or something similar), as it requires basic guidelines and a repository of evaluations to support consistency in impact estimates and in judgments of cost-effectiveness. Indeed, there is no need to wait until projects planned according to the requirements of this approach are reaching completion. Many of the gains from M&E for cost-effectiveness would begin to accrue as soon as significant numbers of completed projects are evaluated according to its specifications. It may be timely, therefore, to sketch out how a professional association dedicated to this approach might function.

The association would have to have a guidebook, criteria for membership, a stamp, a repository, and a standards committee. The guidebook lays out the evaluation approach, stating that each evaluation reconstructs the unfolding of the hypotheses implicit in the project design and estimates project impacts and cost-effectiveness. It would present approaches to estimating impacts and degrees of cost-effectiveness. Any member conducting an evaluation under the association’s stamp would be bound to follow the approach laid out in the guidebook or risk losing membership. This rule is the source of the evaluator’s independence from project management. Each evaluation completed under the association’s stamp would be indexed and included in a database in the repository. The standards committee would be responsible for determining if evaluations comply with the association’s standards.

Monitoring and evaluation for cost-effectiveness aims to increase methodological consistency in order to increase the comparability of impact estimates and judgments of cost-effectiveness, but this should not come at the expense of the rich and valuable methodological pluralism that characterizes the development assistance field. The guidebook should specify the questions to be asked without restricting the approaches that can be taken in answering them. It should give examples of appropriate methodologies and links to evaluations in the database that have used them without restricting evaluators from employing and defending alternatives. Inevitably, there would be disagreements on what constitutes an adequate basis for estimates and judgments. Much of the work in establishing the association would come in reaching a working consensus on these matters, and conflicts would inevitably persist. This kind of conflict is not without costs, but it does support the establishment and maintenance of important powers of judgment among the development community.

Besides supporting an orientation to cost-effectiveness within the development community, such an association would also support methodological advances in evaluation practice. Impact estimates and judgments of cost-effectiveness for which evaluators indicate low or medium confidence would point to areas that could benefit from methodological advances. Within the universe of evaluations carried out by members of the association it would become apparent where assessments of outcomes
are relatively secure and where ambiguity persists. This could help researchers or more experimental practitioners to see where their efforts could productively be applied. It is not unreasonable to expect that the association would create conditions for self-motivated and decentralized improvements in evaluation practice, enhancing methodological consistency as evaluators identify the more effective approaches to impact assessment problems.

There is also no need to wait for the governing bodies of donor agencies to establish weights for combining different kinds of impacts in judgments of cost-effectiveness. An evaluation association could establish these on a provisional basis, and the experience gained through their establishment and application could inform the governing bodies. Just as we would anticipate a decentralized learning process among evaluators on impact assessment methodology, we would anticipate a similar process in regard to judgments of cost-effectiveness. As evaluators observe one another identifying different kinds of impacts as significant for development and presenting different justifications for their valuation, we probably should not expect to see consensus, but we could anticipate the emergence of schools associated with different principles.

VI. LONG TERM GAINS FROM MONITORING AND EVALUATION FOR COST-EFFECTIVENESS

Long term gains from M&E for cost-effectiveness arise from three sources: enhanced accountability, increased learning, and reductions in costs. At present there are some small proportion of projects that from the beginning fail to establish a viable implementation of the project idea. Once funds have been allocated and a team established, management stakeholders have strong interests in the project moving forward. A very small percentage of projects harbor corruption and fraud. Independent and comprehensive evaluations are more likely to identify these so the authorities responsible for criminal investigations can be alerted. More commonly there is a general failure to take the major steps needed to completely redesign or to halt a lame project. If management stakeholders took for granted that the project would eventually be subject to an independent and comprehensive evaluation grounded in the idea of cost-effectiveness, it is much more likely that they would recognize and take account of such shortcomings, and, if necessary, take drastic action. Some project managers and donor agency and government officials responsible for the disposition of donor funds have never learned to construct an organic conception of a development project and its likely impacts. Monitoring and evaluation for cost-effectiveness would lead either to these officials learning to do this or to their being identified by their colleagues and their management responsibilities reduced.

The proposed evaluation approach would also lead more consistently to the recognition of exceptional projects, and to clearer distinctions being made between merely adequate projects and those that are very strong. This would increase the extent to which those professional rewards available to development agencies can be allocated on the basis of contributions to beneficiary well-being.
Monitoring and evaluation for cost-effectiveness increases learning by requiring more broadly synthetic judgments (particularly of cost-effectiveness), by establishing reasons and supports for development professionals to refer to completed evaluations, and by increasing the extent to which evaluations can be compared to one another and provide useful lessons for new programs. I have argued that the act of constructing a mental model of a project, including its evolving relations with beneficiary populations and its interactions with its environment, and of working through the dynamics of the model to make estimates of likely impacts and cost-effectiveness, itself constitutes one of the central forms of learning required for development assistance. To give just one example, for projects that build new institutions this can help managers to anticipate from the start the new responsibilities beneficiary populations will need to undertake. While such projects have often tended to focus initially on more concrete tasks like delivering services and constructing infrastructure, the forward thinking required for estimates of cost-effectiveness can help managers from the start to establish relationships with beneficiaries based on an expectation of the roles beneficiaries will have to play if the project is to succeed. Also, while many managers may have a clear view of the steps that must be taken to carry out the project plan, by grounding their mental model in the idea of cost-effectiveness they can gain a view of appropriate proportions between inputs and outputs. Donor agency and government officials, likewise, can be expected through M&E for cost-effectiveness to gain a better understanding of their strategic choices.

Under the proposed approach impact estimates are likely to be of considerable interest to all management stakeholders. It gives them reason to go to the evaluation association’s database to see what kinds of impacts projects similar to theirs have achieved, and to compare their own strategic situation with those of similar projects. In this way it supports learning between projects. Also, even if evaluations of these similar projects have employed outcome measures and monitoring and evaluation methodologies different from those from their own projects, management stakeholders should still be able to make meaningful comparisons. As long as judgments of cost-effectiveness are made on the same scale, (and assuming that justifications for these judgments are clear and grounded in impact data, and that impact estimates are grounded in an analysis of project strategy,) they should be able to make useful strategic comparisons. The same holds for the development community at large. Whatever the diversity of program strategies and impact assessment methodologies, if judgments of cost-effectiveness are made on the same scale, and assuming that these judgments are properly defended, the development community will be placed in a stronger position to learn from experience and thereby to strengthen program and project portfolios.

Many development professionals are inclined to consider a project successful when it has significant impacts, or when it achieves most of its objectives. It may be that a particular agency has hiring and procurement practices that ensure that costs are reasonable. Perhaps in the absence of cost-benefit benchmarks, professional norms have evolved that simply pay little heed to the cost side. It is clear, however, that in the context of donor agency operations the proportion to costs is a critical factor in interpreting impacts. Otherwise there will be an incentive to pad budgets in one way or another, or simply to pick easy objectives. An important source of gains from M&E for
cost-effectiveness is in establishing an orientation to cost-effectiveness that is sensitive to the cost side. Since donor agencies support operations in countries beset by poverty, and given that their operations are widely observed, it is appropriate for them to err on the side of frugality. Nevertheless when they establish programs that clearly are cost-effective, opportunities for expansion and replication should be aggressively pursued, and the proposed approach encourages this. It should be noted that close attention to costs also tends to favor local hiring and procurement and participatory management approaches that elicit cost-sharing by beneficiaries.

It may be that some of the greatest gains from M&E for cost-effectiveness could come from supporting the more rapid identification of cost-effective investments. By the mid-1990s the basic shapes of trajectories for infections with the HIV/AIDS virus were known, and countries such as Uganda had demonstrated effective containment strategies. Likely mortality and social and economic costs from an AIDS epidemic were, to a broad approximation, understood. Nevertheless, many governments and the donor agencies that supported them continued to promote their established investment strategies. If at that time an orientation to cost-effectiveness had been taken for granted among development professionals, countries such as Botswana, Malawi, Zambia and South Africa, (to name only a few,) might not face the terrible and wrenching crises that the AIDS pandemic is causing today.

Notes

1 The coefficients for the Aid/GDP variable are very small, negative in one case, and with very low t-statistics (0.13 and 0.19 in two different regressions) (Burnside and Dollar, 2000, Table 3, p. 854).
2 Carlsson et al. state categorically in this regard that, “The use of economic analysis in the assessment of aid is subordinate to the organization of the agency, its objectives, decision processes, and incentive structure” (1994, p. 176).
3 Or “apodictic,” the term used by Immanuel Kant.
4 The interesting and important work on valuation in standard economic analysis involves the calculation and use of shadow prices, so that analysis is based on the economic value of costs and benefits rather than the nominal value.
5 Or disability adjusted life years (DALYs), to use the more standard unit.
6 Specifically, the ERR is the discount rate at which the discounted sum of benefits minus costs is equal to zero.
7 An ERR below 0 means that costs exceed benefits in nominal (undiscounted) terms.
In this case rather than updating impact estimates and judgments of cost-effectiveness, evaluators would make original estimates.

References


