

# EVALUATION TO HELP IMPROVE DEVELOPMENT RESULTS

Vinod Thomas<sup>1</sup>  
October 12, 2009

There is growing interest in the independent evaluation function and its application to development contexts. Development evaluators are presented with a rare opportunity to make timely and crucial contributions to development effectiveness. We will need to make use of diverse evaluation methods and lessons of experience to seize this opportunity.

In this connection, this paper addresses a twofold question. First, it deals with the question of the “what” – the need for evaluation methods and institutional frameworks to respond to the growing uncertainties and complexity of development situations. The paper notes the lasting validity of some of the basics in evaluation, and the need to return to the basics. But it also suggests shifts needed to respond to emerging issues, such as growing interlinkages across sectors.

Second, it discusses the “how.” In particular, there are issues of capacity building, incentives, and organization that are necessary for an independent evaluation function to be effective within a country’s institutional architecture.

The paper brings out a number of directions for India to consider in its plan to strengthen the evaluation function. Progress has been made in India in strengthening its monitoring and evaluation (M&E) system. However, several opportunities can be explored to leverage the strategic role that evaluation can play. These areas include using diverse evaluation methodologies; strengthening organizational capacity; ensuring the independence of the evaluation function, with due consideration for engagement; and increasing the demand for high-quality M&E information within government agencies.

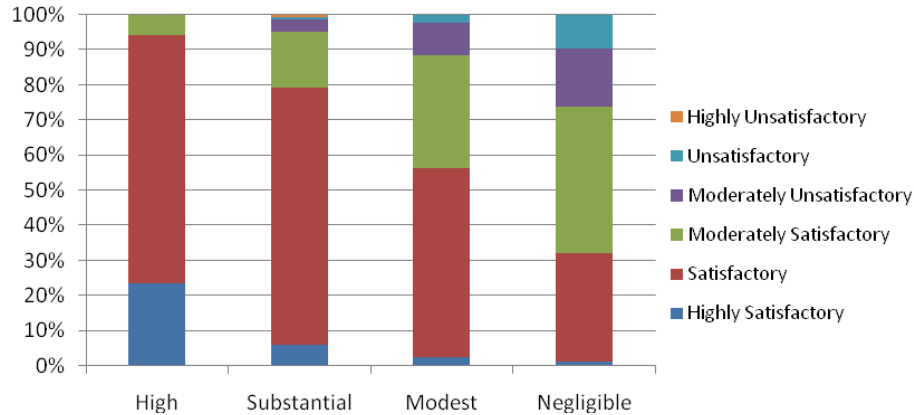
## **1. How evaluations can make a difference in improving results**

Better monitoring and evaluation is associated with better project outcomes. Based on IEG’s evaluations, over 90 percent of the projects that were rated high and about 80 percent of those rated substantial on the quality of their M&E systems between 2007 and 2009 were rated satisfactory or better on their development outcomes. But those with M&E rated modest had satisfactory outcomes or better in only 56 percent of cases. And those with negligible ratings for M&E had satisfactory ratings or better in 32 percent of cases (Figure 1). Whether there is causality between good M&E and better project outcomes is difficult to determine. But these associations strengthen the point that ongoing monitoring information assists with improved management and corrections in program implementation.

---

<sup>1</sup> I am grateful to Jiro Tominaga for working with me on this paper. Special thanks to Nidhi Khattri, Manuel Fernando Castro, Ann Flanagan, Amitava Banerjee and Ximena Fernandez Ordonez for valuable contributions and helpful comments.

**Figure 1. How M&E is related to results in World Bank–financed projects 2007–09**



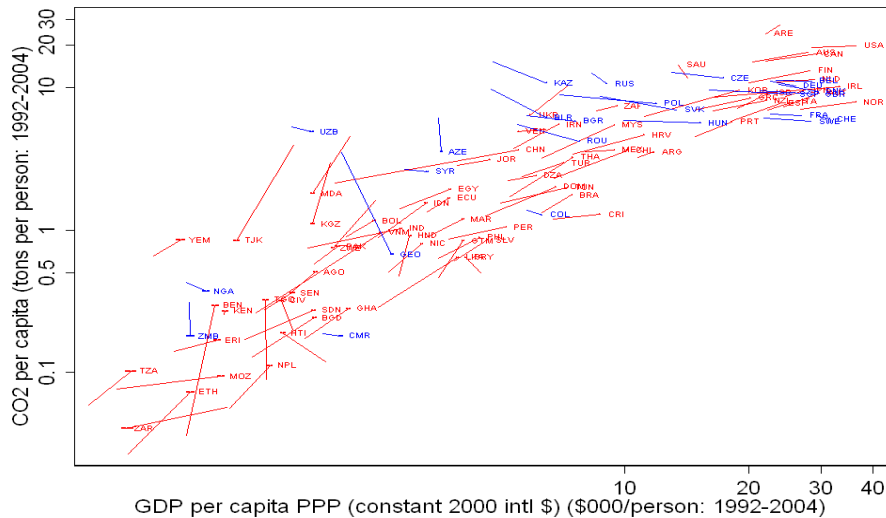
*Source:* World Bank project data and IEG staff calculations.

There is no simple relationship between M&E and growth rates. After all, high and low growth rates have been achieved with good or weak M&E systems. Growth as measured by the increase in gross domestic product (GDP) per capita is critical to poverty reduction, but it masks the differences in the effectiveness with which growth is being achieved or the difference growth makes in welfare outcomes. For example:

- The per capita income in the Philippines was about 30 percent higher than that in Moldova in 2006, but Moldova had a maternal mortality ratio of 16 per 100,000 live births, as opposed to 162 in the Philippines.
- Tanzania and Bangladesh have similar GDP per capita, but the life expectancies of their respective citizens are 55 years and 65 years.
- The data seem to suggest that the maternal mortality rate in India has worsened or, at the very least, has not improved significantly, despite increased public expenditures on health. This rate is six times higher than China's and 14 times higher than Chile's.
- Countries vary significantly in their emissions intensity after adjusting for level of development (Figure 2). The relationship between CO<sub>2</sub> emissions per capita and income per capita remains strong, but it is a band. There can be a sevenfold difference in emissions intensity.

These variations may be due to the initial conditions, but the policy decisions relating to public investment, the regulatory framework, and resource allocations among competing priorities have likely played a role. Continuous M&E of the policy impact is essential to maintain focus and make adjustments along the way, because “[w]hat we measure affects what we do; and if our measurements are flawed, decisions may be distorted” (The Commission on the Measurement of Economic Performance and Social Progress, 2009).<sup>2</sup> Given that government expenditures often constitute more than 30 percent of GDP, if a saving of a small fraction could be achieved while ensuring the same results, the options for policy makers would be significantly broadened. This underscores the importance of rigorous M&E, not only for accountability but also for learning.

<sup>2</sup> The members of the commission are Joseph Stiglitz, Amartya Sen, and Jean Paul Fitoussi.

Figure 2: CO<sub>2</sub> emissions and income

Source: IEG 2009.

Several evaluations have demonstrated the potential to influence the directions of government policies. They need to be methodologically sound, conducted at the right time, focused on key issues of concern at the time, presented in a user-friendly format, and made available to key constituents and beneficiaries. For example:

- A Citizen Report Card evaluation in Bangalore found very low levels of satisfaction with the quality of public service among citizens. It encouraged citizen groups to press for improvements and catalyzed public service agencies to strengthen their customer orientation and improve the quality of services.
- The public expenditure tracking survey conducted in the Uganda primary education sector found that only 13 percent of earmarked funds actually reached schools in 1991–95, and about 20 percent of the funds allocated for teacher salaries went to “ghost workers” who did not exist or were not working as teachers. Government action to publish the intergovernmental transfers of public funds followed, and two locally implemented follow-up surveys showed that the flow of non-wage funds improved from 13 percent to about 80–90 percent reaching the schools. Earlier, most schools did not receive any of these grants, but after the study, 90 percent of them received their full entitlement.

A strong M&E system leads to better governance and improved government accountability. Monitoring information and evaluation findings can contribute to sound governance by enabling evidence-based policy making and by strengthening accountability. A large body of research shows that in the longer term, good governance is associated with robust growth; lower income inequality; reduced child mortality; greater illiteracy; improved country competitiveness and investment climate; and greater resilience of the financial sector.<sup>3</sup>

<sup>3</sup> Governance and accountability are very important in the literature on sustained growth and poverty reduction. A number of empirical studies show the benefits of accountability for the quality of government (Besley and Case 1995; La

## 2. Making sure that basics are in place

Although the importance of M&E systems is widely accepted, quality is far from satisfactory, and opportunities to measure outcomes are often missed. For example, data from World Bank–financed projects closing between July 2006 and June 2008 show that only 37 percent of projects received ratings of high or substantial for the design, implementation, and use of M&E systems. During 1997–2001, half of all approved water supply and sanitation projects cited potential health benefits and close to 90 percent financed infrastructure that very likely would contribute to improved health, but only 23 percent proposed to collect health indicators and only 15 percent ended up collecting them.<sup>4</sup> In the end, 15 percent of projects demonstrated health outcomes.

### *a. Basic principles for good M&E*

A typical problem with an M&E system is lack of baseline data. An evaluation of World Bank support in health, nutrition, and population (HNP) found that a quarter of projects approved in fiscal 2007 had no baseline values in the Project Appraisal Document, and only about half had baseline values for all outcome indicators. Lack of baseline data not only makes it difficult, if not impossible, to assess the effectiveness of activities, but also leads to unrealistic targets and expectations.

Second, selecting the right indicators and defining them clearly are critical for a strong evaluation system. Too often these basics are forgotten at the project design phase, resulting in indicators that are neither relevant nor measurable. This is true not only for projects, but also for government-led development strategies. The other problem is the selection of too many indicators without a clear sense of how to use them effectively for decision making.

Finally, ownership of the institutions involved is essential for successful M&E. An IEG evaluation of M&E in municipal management projects reflects the views of many clients that cities have to consider M&E a useful tool for themselves, not just an instrument for government control or academic use by others. Incentive mechanisms within the institution play a significant role in getting the system established and enforcing it in a desirable way.

### *b. Emerging demand in development evaluations*

These basic principles are valid at all times. The development community is undertaking concerted efforts to deepen the results agenda following the Paris Declaration for Aid Effectiveness (2005) and the Accra Agenda for Action (2008), and achievements have been made in several areas. However, the heightened demand for results and the evolution in the international development architecture pose difficult challenges for evaluators.

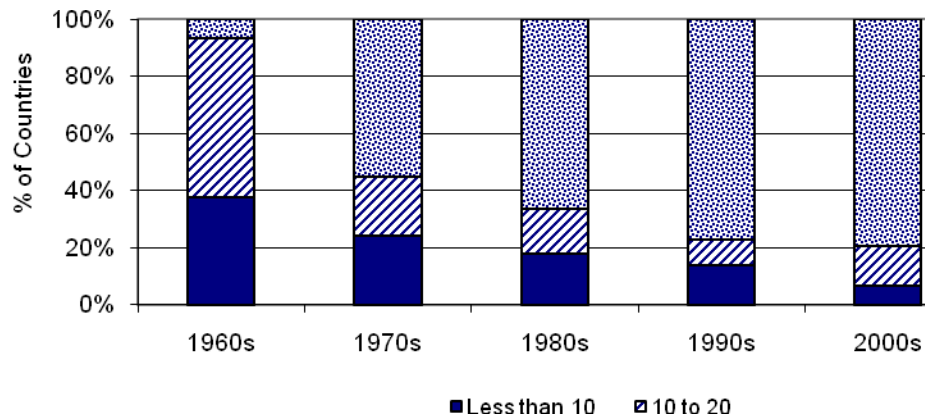
---

Porta et al. 1999; Adsera et al. 2003; Eijffinger and Geraats 2005; Olken 2007; Dyck et al. 2008; Ferraz and Finan 2008; Bjorkman and Svensson 2009; Djankov et al. 2009). Accountability of elected officials is also found to be important for economic growth (Kaufmann et al. 1999; Besley and Burgess 2003). Mauro (1995) finds that the efficiency of the bureaucracy is associated with better rates of investment and growth.

<sup>4</sup> Independent Evaluation Group (2008).

First, the evaluation of development impact needs to accommodate the greater complexity in the development community. Fragmentation in agencies and organizations financing development activities has increased. At last count, there were 280 bilateral agencies, 24 development banks, 40 United Nations agencies, and a plethora of private foundations and nongovernmental organizations (NGOs). Activities around the world are estimated to number 340,000.<sup>5</sup> The number of donors per country tripled, from a dozen in the 1960s to 33 during 2001–05 (Figure 3). Resource delivery has also become more fragmented: there are some 250 major funds or implementing organizations today—100 in health alone. Many of them earmark funds.<sup>6</sup>

**Figure 3. Sources of finance increased dramatically**



*Source:* OECD/DAC – Database on Aid Activities.

Second, further complexity was added when a compact for poverty reduction was reached at the 2000 Millennium Summit. The Millennium Development Goals symbolize the notion of managing for development results, which implies extended causality chains of interlinked activities and interventions.

Third, despite the growing difficulty of defining attribution, scrutiny to ensure accountability of development aid is more intense than ever before. Sharpening evaluation methods can help in this regard, but it is becoming increasingly realistic to talk of contribution rather than attribution.

Fourth, development evaluations have to examine the vital role of the private sector in development. Of significance is the synergy and missed opportunities of the interactions between public policy and private investments.

Finally, the recent global financial crisis put a heightened premium on the speed with which evaluative findings could be utilized. While this development offers an opportunity, it also tests the adaptive capacity of evaluators to the limit. The challenge is not only to embrace the flexibility to respond rapidly to the need, but also to develop the evaluation methodologies that can offer guidance as crisis response measures are being prepared.

<sup>5</sup> Eckhard Deutscher and Sara Fyson. (2008).

<sup>6</sup> International Development Association. (February 2007).

### 3. Evaluating linkages and synergies

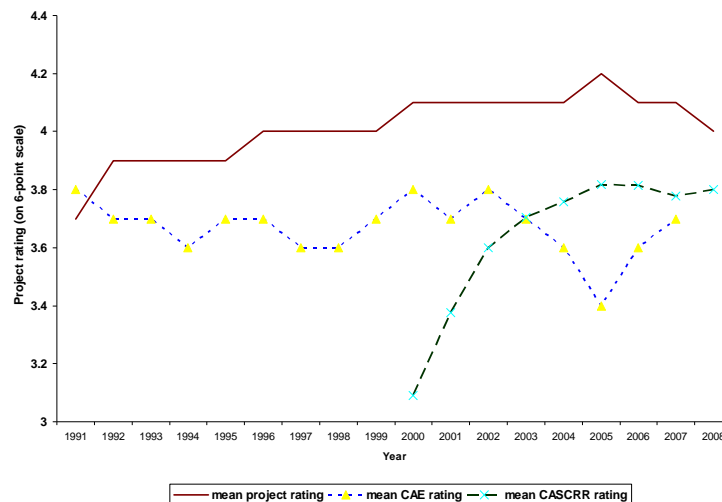
Given recent developments, evaluations need to capture cross-project links. A project's objective might be stated as providing clean water, but the evaluation would be flawed if complementary health benefits or resettlement costs were omitted or if unanticipated effects on the city's growth and incomes were not considered. In IEG's trade evaluation, we found that the objective was often to liberalize trade, but results on exports depended on infrastructure and institutions that the projects did not target.

#### *a. Disconnect between country programs and project ratings*

Country programs have become a platform for assistance strategies for many donor institutions. It has therefore become increasingly common to assess the development effectiveness at the country-program level. IEG provides two vehicles for ratings of the World Bank's overall programs for specific countries: Country Assistance Evaluations (CAEs) and Country Assistance Strategy Completion Report (CASCR) Reviews.

Given that the expected outcomes of country programs differ from those of projects, there is no presumption that the ratings should be the same. However, there is an expected association among country performance, country programs, and projects. A recent IEG review shows that the country ratings tend to be lower than project ratings (Figure 4).<sup>7</sup>

**Figure 4: Disconnect between project performance and country outcomes**



*Source:* World Bank database and IEG calculations.

*Note:* To compute the country program rating by fiscal year, the relevant Country Assistance Evaluation (CAE) and Country Assistance Strategy Completion Report (CASCR) Review ratings are applied to each year covered by the evaluation. Poorer results in earlier CASCR Reviews may indicate a steep learning curve for the Bank in using the CASCR as an instrument (it was introduced in fiscal year 2003).

<sup>7</sup> IEG (forthcoming)

Country program outcomes are measured against country objectives such as growth, poverty reduction, and environmental stewardship. The broader the objectives are, the greater the number and scale of factors that can affect outcomes—for example, the policy and institutional content and implementation of the country program, as well as exogenous factors including shocks to the system. The underlying reasons point to the tougher challenges of country program success.

***b. Shift to results framework***

A challenge is to construct a results framework of the country program, with projects and programs as integral components of the results chain. Projects must be relevant to a country's strategy and synergies between projects must be captured, while keeping in mind the role of external factors.

Most evaluation groups in multilateral development banks (MDBs), including IEG, use an objectives-based evaluation framework rather than a fully results-based framework. While in either case objectives may be set out early in the program, in the objectives-based approach, the inquiry is whether the stated objectives were met, whereas in the results-based model the emphasis is on achieving outcomes and impacts. For example, consider a program that promotes alternative dispute resolution by educating the public on the benefits of mediation centers and expert mediators. The outputs would refer to the number of radio announcements on mediation, or even the number of people accessing mediation centers. The outcome would refer to a decrease in backlogs of court cases or cheaper access to justice for individuals. The impact would be the development of good governance.<sup>8</sup> Two key attributes of strong M&E systems are essential for the shift from objectives-based system to a results framework: baseline data and measurable outcome indicators.

A second way to get closer to results is to consider complementary, unanticipated, or omitted objectives. These complementary or conflicting aspects have often turned out to be important in IEG's evaluations. For example:

- In primary education, we found success in greater access for children, but also that learning outcomes were not targeted.
- We found that the CAS had missed focusing on natural resource management in Indonesia; governance and the private sector were under-emphasized in Ethiopia; and the strategy did not focus adequately on poverty in Egypt and Georgia

Third, evaluations can better formulate why changes would be expected. This broader focus on results is strengthened if logic about how the intervention is expected to bring changes is embedded in the approach. This relates to how we believe individual, intergroup, and societal change happens, and precisely how interventions will produce positive results.<sup>9</sup>

Behind the set of expectations are assumptions. For example, a micro-enterprise program for poor women might be expected to help generate profits that would go into the business and help children's education. The program, however, might be assuming that the enterprise is marketable,

---

<sup>8</sup> Thomas (2008).

<sup>9</sup> Shapiro, I. (2005).

that husbands would be supportive, that women would have control over profits and use them to grow the business and for children's education, and that children would not be called on to do even more household work.<sup>10</sup> The evaluations need to assess whether the program has captured the essential assumptions and the theory of change in achieving results.

Finally, there is a growing need for evaluations to capture the impact of public-private coordination. Evaluation has shown there is scope for greater collaboration between the World Bank and IFC and MIGA.

In evaluating the private sector led development projects, some MDBs and bilaterals attempt to measure additionality. The Asian Development Bank asks whether its financing was necessary for the timely realization of the project. The Inter-American Development Bank defines additionality as the value added by its contribution to enhancing a project's long range sustainability prospects or its development benefits.<sup>11</sup>

But project-level additionality is not enough to capture the synergies created in public-private partnerships. The analysis could be strengthened by measuring additionality at the project, program, and systemic levels. Moreover, it is time to create an effective link between the evaluation methodologies for the public and private sector activities under an integrated framework. IEG is in a unique position to leverage the expertise in evaluating the public sector (IEG-World Bank) and private sector led activities (IEG-IFC and IEG-MIGA). In fact, IEG pioneered a first review of the additionality of IFC's operations in its FY2008 Annual Report on Development Results.

#### 4. Impact evaluations and attribution

The need to assess the counterfactuals is becoming increasingly important. Impact evaluations to clarify attribution at the project level are useful tools for understanding the links between the project and outcomes. They can also show that strong development impacts come from interactions across sectors, as shown in several examples below.

- Impact evaluations supported the effectiveness of conditional cash transfer programs in fostering school attendance and children's visits to clinics at a time when these programs were being dismissed by development practitioners.<sup>12</sup>
- Impact evaluation of the Training and Visit extension system in Kenyan agriculture found the system to have no impact on agricultural production, even though it had been promoted as the standard for extension services.

Another benefit of rigorous impact evaluations is that they can help depoliticize decision making by focusing support for programs on empirical evidence of benefits, as in the case of the Mexican *Oportunidades* program.

---

<sup>10</sup> Thomas (2008).

<sup>11</sup> IEG (2008c).

<sup>12</sup> Gertler, Paul (2004).

*a. Essential factors for useful impact evaluations*

Several factors contribute to the usefulness of impact evaluation. Impact evaluations need to be demand-driven and not a passing fad. The early 1980s saw a similar focus leading to the establishment of a department within the World Bank devoted to M&E just for the agricultural sector. Much high-quality work was done, but the department closed after a few years and the quality of M&E in agricultural projects has almost certainly declined as a result, chiefly because M&E systems were over-designed.

Impact evaluation needs to be timely. Lack of timely impact evaluation data can result in the continuation of bad programs. For example, a World Bank nutrition project in Bangladesh was scaled up before a thorough evaluation of a smaller-scale version of the project had been carried out. Once an evaluation was conducted, the project was found to have had limited impact, and several design problems were identified.

There are gains to doing impact evaluations ex ante rather than ex post. Building in impact evaluations early also can encourage better project design and real-time feedback. Poor institutional memory can make it difficult to trace results ex post. It is more practical to do the evaluation while the project is still open. But there are also challenges to building in ex ante evaluations.

There is a tendency to measure things that are easy to measure with readily available tools. Of over 200 World Bank impact evaluations already catalogued, more than half are in social protection, many for cash transfers. Just 2 percent are for agriculture, showing that there are clear gaps in coverage. Little progress has been made on impact evaluation of programs aimed at improving the environment. We need to look beyond what's visible under the lamppost.

The design of impact evaluation needs the foundation of development experience. It pays to examine the causal chain from inputs to outcomes. IEG reanalyzed data from a nutrition project in Bangladesh, using a national survey to construct a control group through propensity-score matching, since the project survey control group was weak. Very little impact was found because nutritional counseling was targeted to mothers, whereas men do the food shopping in rural Bangladesh. Furthermore, the mother-in-law usually has the final say on many matters relating to childcare and diet, not the mother. So messages need to reach a broader audience, and social workers need to be trained accordingly.

Some might argue for experimental designs with random assignment to a treatment group and a control group. These evidence-based evaluations are often called randomized control trials (RCTs). An issue with RCTs is that the measured impacts will very likely vary with population characteristics when a project is scaled up. Most often a scaled-up project will provide services to a much poorer population than the experimental group. In addition, results may be heterogeneous among the population, making the mean effect less relevant.<sup>13</sup> Impact evaluations are defined by a counterfactual, not necessarily by randomized controls, which can be achieved using quasi-experimental designs that use a comparison group, but without necessarily having random assignment. While there are (at least) two schools of thought on RCTs, one estimate considers only

---

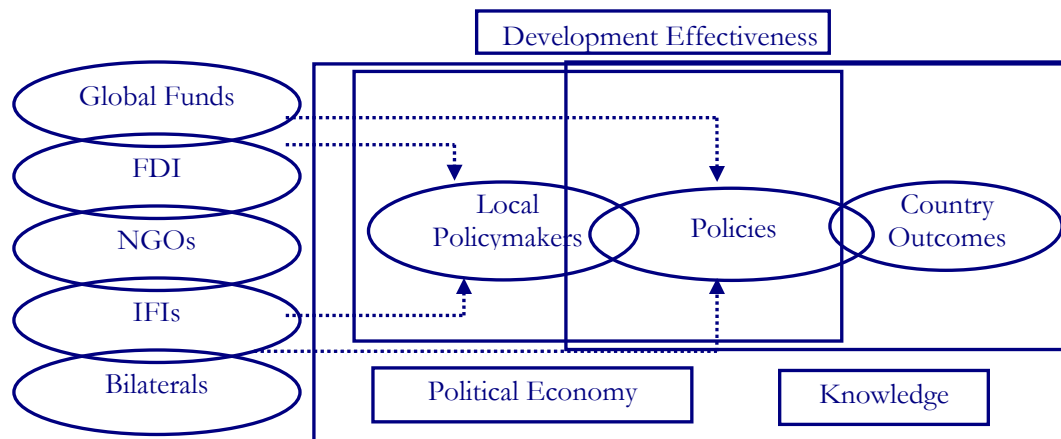
<sup>13</sup> Ravallion, Martin (2008).

about 5 percent of the total value of development finance—at most—amenable to such an approach.<sup>14</sup> If this is true, careful selection is required to build a collection of policy-relevant knowledge.

***b. Defining attributions***

Impact evaluation is a powerful tool, but defining clear attribution for many development projects remains difficult. There are difficulties with making a case for attribution at the project level, even when experimental designs are used. Consider now the difficulties in attributing impact at the sector or county level, and how problematic it would be to measure the partial effects of one donor at the country level (Figure 4).

**Figure 5. Attribution is not easy with multiple players**



*Source:* IEG staff.

External factors, beyond the control of any of the players considered, also matter. For instance, institutional and policy progress made in Europe and Central Asia in countries taking loans for public sector reforms was similar to progress in those that did not (Table 1). A common external factor explaining this performance seems to be the requirements for accession to EU.

<sup>14</sup> Bamberger, M., and White, H. (2007).

**Table 1. Policy environment improves, with and without public sector reform**

Public Sector Reform Loan	Yes		No	
Africa	70%	(30)	47%	(15)
East Asia and Pacific	70%	(10)	56%	(9)
Europe and Central Asia	90%	(20)	86%	(7)
Latin America and Caribbean	75%	(20)	25%	(8)
Middle East and North Africa	57%	(7)	0%	(2)
South Asia	50%	(6)	0%	(1)
Total	73%	(93)	48%	(42)

*Source:* WB CPIA scores and IEG staff calculations

*Note:* Total number of countries in parentheses.

Attributing effectiveness when there are multiple players is tough at the country level, let alone at the global level. And yet, getting a handle on these larger impacts is increasingly important to ensure that complementarities are exploited and that conflicts among the work of different players are minimized. This is all the more important with the growing numbers of development partners and the highly globalized setting within which we operate today.

Global public goods and globalization increase the attribution challenge. More proactive steps toward global and regional coordination are being called for to address rapidly emerging cross-border problems such as climate change, communicable diseases, international trade agreements, and other supranational issues. Many important global public goods need collective action across countries to be delivered effectively.

Given the difficulty, contribution, rather than attribution, may need to be assessed. Contribution Theory<sup>15</sup> suggests that knowing the results chain is enough to make a plausible case that a program contributed to outcomes, having explored alternate explanations. Some argue that when it comes to country evaluations, we should be looking at how the country or sector has done overall, and indicate how the donor organization has likely contributed, along with many others, to the country's positive and/or negative outcomes.<sup>16</sup>

## 5. Institutional framework for strategic evaluation

Evaluations are conducted to provide findings and recommendations relevant to the strategic directions of the evaluated organizations. Within the World Bank Group institutional architecture, IEG aims to develop findings that can shape the Bank Group's directions to achieve the greatest development impact. The quality of the evaluations based on rigorous methodologies adapted to a given situation is critical.

### *a. Strategic directions for evaluation groups amid global uncertainties*

The global uncertainties we are experiencing add new dimensions to the strategy of evaluation groups. It has become increasingly important to have the flexibility to shift the evaluation program

<sup>15</sup> Mayne, J. (2001).

<sup>16</sup> Picciotto, R. (2007).

as needed. Two attributes have particular significance in this regard: selecting the right evaluation topic and providing timely inputs.

Evaluation topics need to be selected keeping in mind the strategic issues of the institutions and associated trade-offs. And in a rapidly changing environment, the strategic directions of an institution shift constantly. There are also emerging issues that can be expected to pose challenges to the existing business model in the medium term. A case in point is the climate change agenda, which has implications across sector and country boundaries. Anything less than global solutions could imply failure, but the evaluations based on programs at the country level may not be useful for improving outcomes at the regional or global level. Other issues such as trade, the financial system, environment, and water resource management would also require cross-boundary cooperation.

The impact of evaluation is substantially enhanced when the findings are presented at the right time. This is particularly important in a time of uncertainties when responses need to be developed in a limited timeframe. Not leveraging the knowledge of similar events in the past is a missed opportunity, and evaluators can make considerable contributions in this area. Timely inputs by IEG on the experiences of past financial crises offered useful guidance in developing the response package of the World Bank Group to the recent financial crisis.

Selecting the right evaluation topics and providing timely inputs require deep understanding of the institutions which can be gained through close engagement. The independent evaluation functions in international financial institutions (IFIs) are maintained within the architecture of the respective institutions in order to achieve the dual roles of independence and engagement. However, the design encompasses an inherent source of tension between the two roles.

#### ***b. Organizational and behavioral independence while ensuring engagement***

Independence is essential in ensuring objective accounts of the institutions' effectiveness. Without independence, evaluative findings could lose credibility, and thus their impact. An independent evaluation function is also important to prevent gaming of internal incentives and disincentives. There are two aspects which have to be in place for independent evaluations to be carried out.

First, the mandate and institutional processes must clearly recognize the independence of the evaluation function. Key components include the processes for appointment and evaluation of the head of the evaluation group. In the case of IEG, the Director-General, Evaluation (DGE) is appointed by the Executive Directors of the World Bank Group at the level of Senior Vice President. The DGE is not eligible for subsequent appointments in the World Bank Group. The DGE reports to the Board of Executive Directors; not to World Bank Group management. IEG reports its findings to the Board without management clearance, though it allows a management response. IEG sets its own work program and budget in line with its strategic directions and the Bank Group's priorities. The Board endorses IEG's work program and approves its budget.

IEG is also free to decide on evaluation topics and their timing; mount evaluation missions; and consult with NGOs, governments, the private sector, and all other sources as needed without prior clearance from Bank management. There is disclosure of evaluation text without negotiation with the Board or the management.

Equally important is behavioral independence. Appearance matters, and organizational independence must be enhanced by behavioral independence. In particular, the issues involved with the potential conflict of interests need to be addressed systematically. It is common to allow evaluators to move to the rest of the parent organization. Controlling the hiring and promotion processes could be an effective way to manage the risks. A complementary measure taken in some evaluation groups, including IEG, is to develop and issue the recusal guidelines.

Finally, it requires determination on the part of the management team of the evaluation group to protect independence. An important task of the group is to deliver hard messages when needed. The group is constantly under pressure to soften the message. To ensure credibility, tough decisions have to be made in favor of clearer messages as opposed to self-restriction, keeping in mind that the accountability of the evaluatees will be strengthened that way.

*c. Incentives for a strong monitoring and evaluation system*

Another critical element is the role of a complex set of incentives to produce and use accurate information. How effectively the system functions—how accurately the data are collected, how robustly the data are analyzed, and the degree to which performance information is shared publicly and widely—is largely dependent on the ways in which the information is used.

Demand for specific information on outcomes and impacts plays a key role in extending measurement to those levels and keeping the system honest. For example, parents' access to data on school performance is a powerful motivator of performance for school administrators. It should also be the motivator for generating the performance data in ways that are useful, not only for accountability but also for learning, improvement, and rewards reflected in additional resources. And once accurate performance system is available, it, in turn, generates further demand for M&E information. With the demand for outcomes-based performance systems, the debate shifts from inputs and activities to outcomes of interest to the beneficiary. Beneficiary-led demand can come in the shape of loosely organized interest groups (e.g., parent associations) or civil society organizations.

Higher-level management needs to recognize the importance of performance information and to provide incentives for objective assessment of its activities. For a strong monitoring function, the demand is generated from within. The key is to link the contributions to M&E at the appropriate level to individual performance, which remains as a challenge for many development institutions, including the World Bank. Independent evaluations provide an external incentive to produce and use accurate performance information. As for government systems, different countries have attempted to address these incentives-related challenges to varying degrees.

- The U.K. improved its performance management system by articulating clear objectives and targets for public service, communicating these aspects to the public-service providers, and introducing the public service agreements with clearly measurable targets. The objectives are based on outcomes that matter to the public, and service providers explain what they can deliver in return for additional investments. The Treasury monitors progress, commissions evaluations, and prepares public expenditure reviews.
- Mexico's public management reform started in 1997, but took root only after the successful evaluation of Progresa. In 2004, CONEVAL was established to regulate the government

M&E system, which set out results frameworks, indicators, and a system of evaluations. The system was strengthened by introducing performance-based budgeting in 2007, which significantly enhanced the incentive for performance.

- Similarly, in Colombia, the National Results-Based Management and Evaluation System (SINERGIA) has established performance indicators and set out an ambitious agenda for impact evaluations. The system actively engages civil society in disseminating its reports. Information is reported publicly, used within the government to assess agency/organization performance, external evaluations are used to obtain independent information, and there are efforts to strengthen performance-based budgeting.

In other cases, M&E systems have been designed and instituted through regulatory mechanisms, but these systems have languished due to weak incentives to generate and use performance information: lack of demand from constituents (e.g., parliamentary bodies and civil society); from within the system (e.g., monitoring focused on inputs, disconnect between budgeting and performance); and within the organization (complete disconnect between personnel performance and appraisal).

#### *d. Organizational capacity as the foundation for strong M&E*

In addition to incentives, organizational capacity and public officials' skills and knowledge play a vitally important role in an effective country evaluation system. Institutional arrangements are not enough; there must also be the organizational capabilities, skills, and knowledge to carry out the evaluations and to use performance data for management and learning and to continue to generate desired organizational performance. There are several key aspects to institutionalizing effective evaluation systems.

- Effective systems rely on regulatory and public sector reforms that explicitly incorporate accountability and use of performance information (outcomes and impacts) for program design and policy making.
- Second, the backbone of such systems is the information infrastructure—regular data and statistics.
- Third, the use of information requires knowledge and skills among not only technical staff, but also key public officials managing the system.
- Finally, an underlying feature of all of these aspects is the willingness in government to experiment and take a systematic trial-and-error approach. No one system is perfect for all circumstances, and none will be perfect forever for a given circumstance.

Such capacities, knowledge, and skills may be generated in several ways. First organizational capacities can be enhanced by streamlining functions and clarifying organizational goals, mandates, and activities. These steps involve, for example, reviewing and eliminating duplication and overspending and consolidating functions (e.g., data collection on key indicators); instituting mechanisms for interagency collaboration on information and data (e.g., how line ministries can work with the statistical agencies for timely information); and raising the awareness of the key managers in the organization of importance of performance information.

Second, the gap in knowledge and skills can result in poorly generated, analyzed, and reported data. The individuals involved in the M&E system must have basic competencies. The challenge is massive, and indeed apparent in almost all systems, where data are often weak and poorly analyzed,

raising questions about true performance. Weakness of monitoring systems due to lack of capacity is often identified in IEG evaluations. This issue pertains to a broader challenge of how a system recruits and retains its staff, but it is also relevant for how the system maintains and renews the skills of its current staff, based on specific organizational needs.

IEG has been addressing this issue through its annual program, the International Program for Development Evaluation Training (IPDET). IEG plays a catalytic role in enhancing M&E capacity and is now beginning a program with several donors, partner countries, and institutions to help strengthen institutions based in partner countries provide demand-driven, cost-effective M&E capacity building services. The program initially will support four institutions (or consortium of institutions) to develop curriculum, advisory services, and technical assistance in response to specific regional needs and demands. The institutions will be selected competitively in the next few months.

## **6. Going forward for India**

The development evaluation community is being propelled into a world of added complexity and uncertainties. Results measurement has gained a prominent status in the new development paradigm, but the extended results chain makes it challenging to establish simple links between results and interventions. Maintaining the independence of the evaluation function is essential, not only for the credibility and accountability of the institution, but also for learning and responding rapidly to a quickly changing world. But a rapidly shifting environment requires increased engagement with other partners and evaluatees for evaluators to stay relevant in influencing strategic directions. In a time of uncertainty, clear opportunities exist to provide guidance for the future based on the analysis of the past as well as the present. To make a difference and to be influential, evaluations need to go beyond monitoring of stated goals and seek to capture full impacts as well as lessons in implementation.

### *a. What M&E systems and approaches offer to India*

India has made strides in its M&E institutional framework and methods – in the “what” of its M&E system. An institutional framework for M&E is in place and it has many of the components for strong M&E functions seen in Organisation for Economic Co-operation and Development (OECD) and middle-income countries, such as U.K., Mexico, Brazil, Colombia, and Chile. It has an independent Programme Evaluation Organization (PEO) that undertakes external evaluations at the request of the Planning Commission, ministries, and departments. The Development Evaluation Advisory Committee (DEAC) has been established, with members drawn from e academia and the administration. DEAC identifies and prioritizes programs to be evaluated, ensures links of the PEO and agencies involved in the evaluation process; provides guidance on national evaluation policy, and develops strategies for evaluation capacity development. At the subnational level, the State Evaluation Organizations (SEOs) are the counterpart of the PEO and, with significant differences in capacities, are in charge of evaluating most state-level programs. Some states conduct around 20 evaluation studies each year, while most conduct only 1 or 2.

An increased focus on results, and results measurement, could make a key difference in projects and policy implementation. India’s growth rate has been substantial; yet the challenges remain enormous.

India's human development index is ranked at 132 out of 179 countries (UNDP Statistical Update, 2008). The incidence of poverty, although declining, remains high (22–28 percent in 2004–05),<sup>17</sup> with malnutrition as high as 45 percent, as measured by underweight children under three years of age.<sup>18</sup> Both figures exhibit large variations by state. Immunization rates for Brazil and China are over 90 percent, while for India they remain at 55 percent (2005–07),<sup>19</sup> and life expectancy at birth is 65 years (2007), while Brazil and China have each achieved life expectancy of over 70 years. Most environmental indicators show negative trends, with potentially huge negative impacts on sustainable growth.<sup>20</sup> Service delivery in health and education remain challenges, especially for the poor.

Given the country's great potential but huge challenges in finding the right approaches, what strategic roles can M&E play in India?

First, evaluation systems must measure and bring attention to key outcomes and the factors and interventions that lead to those outcomes. For example, small-scale evaluations done in Uttar Pradesh in the context of Sarva Shiksha Abhiyan (SSA) showed that higher learning outcomes were achieved when community volunteers were taught remedial reading approaches for illiterate children. Interventions designed to enhance community participation in and of themselves had limited effects (Banerji et. al. 2008). Despite growing rates of enrollment, a nationwide rural survey of 7–14 year-olds found that 35 percent of school-age children could not read a simple paragraph and 41 percent could not do simple subtraction (Pratham 2005). How effective program interventions are in achieving not just immediate outputs, but key outcomes and longer-term results are a clear focus of any M&E system.

Second, a streamlined and strategic evaluation function helps identify early on the most effective strategies to attain desired outcomes, given investment constraints. Evaluations that place detailed focus on program design and delivery mechanisms help in making mid-stream corrections or decisions regarding large outlays. For example, Duflo and Kremer's (2003) evaluation in Kenya showed that to keep students in school, de-worming cost only about \$3.50 per child per year. In comparison, estimates showed that school-feeding cost \$36.00 per student per year, and school uniforms even more (Duflo, E. & M. Kremer 2003). Such knowledge regarding relative effectiveness becomes available through systematically comparing different approaches to attain the same outcomes. This approach is particularly useful for complex interventions requiring inter-sectoral linkages. GoI's large-scale schemes such as NREGA, SSA, and Mid-Day Meals (which have decentralized implementation models) provide a rich set of materials for not only evaluating impact, but also for assessing issues of implementation, efficiency, and cost-effectiveness. Lessons from such schemes would provide rich material for learning and improvement of program implementation.

Third, evaluations can identify bottlenecks in effective use of resources. Achieving India's ambitious economic agenda and Millennium Development Goals may require additional efforts in identifying not only the major drivers of such outcomes, but also the causes of implementation failure. Research

---

<sup>17</sup> India's Planning Commission.

<sup>18</sup> National Family Health Survey (2005-06).

<sup>19</sup> World Bank Development Indicators.

<sup>20</sup> Country Assistance Strategy (2000-2012).

has shown, for example, that medical staff at primary health centers in India had absence rates of 40 percent and teachers in primary schools had absence rates of 25 percent (Chaudhury et al. 2006). Similarly, PETS (public expenditure tracking surveys) have identified not failures of specific interventions, but rather resources not being utilized for the purposes for which they were designated.

The evaluation system in India thus may need to consider not just what should be evaluated, but an overall strategy of how and when the evaluations should be conducted and what methodologies could be used for different purposes. In today's dynamic environment, evaluators need to respond rapidly to emerging issues. Evaluative findings can make considerable contributions as the responses to unanticipated events are being developed, such as in the aftermath of the Tsunami.

Methodologies that go beyond ex-post evaluation to enable prospective evaluations are an important tool in a comprehensive M&E system. Mixed-method approaches hold promise in assessing project interventions and their impacts, as do more rigorous quantitative impact evaluations. Even though there seems to be a decent amount of ongoing impact evaluation work, strategic selection of topics and timing is essential to influence policy debates. Establishing how different types of evaluations can be instrumental for different purposes is therefore one point of consideration in India.<sup>21</sup>

A challenge with growing significance is to look across both public and private sectors to provide a broader view of development impact in a country including large scale private sector activities. This issue has particular relevance to India as large financial institutions and banks finance private sector projects that may have significant development and environmental impacts. Financial institutions such as the State Bank of India, ICICI Bank, IDBI Bank, IDFC, and the India Infrastructure Finance Company clearly finance large scale investments that affect several lives.

***b. How India can continue its strategic directions in M&E***

In addition to the institutional setting, strong organizational capabilities, and personnel competencies provide the basic foundation for an effective M&E system. An effective system draws on a wide array of evaluation approaches, tools, and methods to respond to the heightened demand for reliable, high-quality information. Thus, consideration needs to be given to how these aspects can be introduced through in-training programs for officers, especially those in State Evaluation Offices and field units. Capacity for technically demanding aspects of M&E (e.g., impact evaluation) is low both within government and in the institutions that typically conduct evaluations for government. Publicly funded institutions supported by the Indian Council for Social Science Research have indicated an interest in finding out about the benefits of impact evaluations in government and how to design and conduct them. Systematic investments in developing the M&E capacities of stakeholders (e.g., government agencies, academia, the private sector, civil society, the media, etc.) will enhance the effectiveness of the M&E system

Ensuring organizational and behavioral independence of the evaluation function is crucial for the credibility of evaluations. India's evaluation model focuses on independence and makes important use of external academic institutions. However, ensuring high-quality evaluations requires

---

<sup>21</sup> Informal interviews with government officials.

knowledge of the strategic issues that can only be obtained through engagement with the evaluatee. An example of such engagement is the approach taken by the World Bank's Social Protection team in India; the team invested in raising awareness among its state counterparts. Evaluatees found value in the kinds of policy questions that the evaluations would inform and understood how these could improve their programs. This resulted in an increase in the demand for a number of evaluations, including those of the National Rural Employment Guarantee Act (NREGA) and the Public Distributions System in Bihar. There is a need to find an appropriate modality for engagement with the evaluatees on design while ensuring independence of the evaluation. A strong emphasis on dissemination of the findings is also encouraged to expand the reach of these evaluations.

The tension between independence and engagement, however, needs to be managed in a systematic and transparent manner. As part of this process, a transparent discussion of data and findings with beneficiaries and civil society might also help increase attention and credibility for evaluations and enhance their learning and problem-solving functions.

Good performance information remains a challenge in India. Although a robust statistical system is in place, with defined responsibilities for the collection of data, timeliness and quality of basic monitoring data remain an issue with users, especially with respect to results. There seems to be more interest in monitoring expenditures and outputs (e.g., kilometers of roads built) than in measuring the results associated with those roads, even in terms of intermediate outcomes such as increase in usage, reduction in travel time, or reduced transportation costs. In the words of a high-level official interviewed for recent World Bank research: "my job is to build roads...use is immaterial," and feedback from several government officials indicates that the emphasis on expenditures outweighs the emphasis on outcomes. Strengthening this will require coordination across planning and statistics agencies, as well as between the central and state levels.

Closer attention is likely needed to evaluate the activities based on public-private partnerships (PPPs). The areas particularly relevant in the Indian context are the extractive industries, agribusinesses, and infrastructure, where significant private sector investments are likely to be mobilized to meet domestic needs. For examples, the financing of two mega coal based power plants in India—Sasan and Tata Mundra, or the activities of large agri-based companies, such as ITC are financed by several financial institutions that could adopt an evaluation methodology to assess the development impact of the private sector projects they finance. The point here is that these private sector investments are often large in scale and have significant development impacts that need to be monitored. As PPP is an important strategic direction for the Government of India, evaluations should be prepared to provide timely inputs to guide the process. Attempts are being made in IEG to address this challenge by evaluating the issues which cut across the World Bank (IBRD/IDA), IFC, and MIGA with the team with expertise in the public and private sector activities. Recent examples include the evaluation of the guarantee instruments of the World Bank Group, and the World Bank Group support for environmental sustainability.

An increased demand from government officials for high-quality M&E is essential; utilization of M&E findings and recommendations is directly related to the program officials' appropriation of the evaluation findings. But increasing the demand for M&E has implications for changes in the overall incentive structure. Growing public demand for accountability creates a clear incentive. A key achievement in this regard was the enactment of the Right to Information Act in 2005, which enhanced access to information in the public interest. This was a major policy shift toward greater transparency and has supported public demand for greater accountability for results. In addition,

institutional mechanisms to reward high-quality evaluations and effective use of M&E work will also boost demand.

M&E has incisive effects when it is used for decision making. Thus, clear links between results, planning, and budgeting need to be considered. For good performance information to be generated, robust performance management systems at the agency level may need to be developed and institutionalized. Shifting attention from outlays and physical targets to real outcomes in decision making is a challenge in all systems, but one that has enormous pay-offs for development effectiveness.

## References:

- Adsera, Alicia, Carles Boix, and Mark Paine. 2003. "Are you Being Served? Political Accountability and Quality of Government." *Journal of Law, Economics, & Organization* 19(2): 445-490.
- Bamberger, M., & White, H. (2007). "Using strong evaluation designs in developing countries: experience and challenges" *Journal of Multidisciplinary Evaluation* 4(8): 58-73.
- Banerji, A.V., R. Banerji, E. Duflo, R. Glennerster, S. Khemani (2008). "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." Policy Research Working Paper, 4584. The World Bank.
- Besley, Timothy, and Anne Case. 1995. "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits." *Quarterly Journal of Economics* 110(3): 769-798.
- Besley, Timothy, and Robin Burgess. 2003. "Halving Global Poverty." *The Journal of Economic Perspectives* 17(3): 3-22.
- Bjorkman, Martina, and Svensson Jakob. 2009. "Power to the People: Evidence from a Randomized Field Experiment of Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* (forthcoming).
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers (2006). "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives* 20(1), 91-116.
- Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2009. "Disclosure by Politicians." NBER Working Paper No. 14703.
- Duflo, E. & M. Kremer, (2003). Use of randomization in the Evaluation of Development Effectiveness. Paper prepared for the World Bank's Operations Evaluation Department.
- Eckhard Deutscher and Sara Fyson. 2008. "Improving the Effectiveness of Aid." *Finance and Development* v. 45, n.3. Washington, D.C.; International Monetary Fund
- Eijffinger, Sylvester, and Petra Geraats. 2005. "How Transparent are Central Banks?" *European Journal of Political Economy* 22(1): 1-21.
- Ferraz, Claudio, and Federico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123(2): 703-746.
- Gertler, Paul. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." *Health, Healthcare and Economic Development* 94(2).

Independent Evaluation Group. 2008a. *The Health Benefits of Water Supply and Sanitation Projects: A review of the World Bank Lending Portfolio*. IEG Technical Working Paper. Washington, D.C.: The World Bank.

----- 2008b. *Climate Change and the World Bank Group Phase I: An Evaluation of World Bank Win-Win Energy Policy Reforms*. Washington, D.C.: The World Bank.

----- 2008c. *Independent Evaluation of IFC's Development Results 2008: IFC's Additionality in Supporting Private Sector Development*. Washington, DC.: International Finance Cooperation

----- forthcoming. *The World Bank's Country Policy and Institutional Assessment (CPIA)*. Washington, D.C.: The World Bank.

----- forthcoming. *Annual Review of Development Effectiveness 2009, Achieving Sustainable Development*. Washington, D.C.: The World Bank.

International Development Association. February 2007. *"Aid Architecture: An Overview of the Main Trends in Official Development Assistance Flows."* Washington, D.C.: The World Bank.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 1999. "Governance Matters." Policy Research Working Paper No. 2196. Washington, D.C.: The World Bank.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1999. "The Quality of Government." *Journal of Law, Economics and Organization* 15: 222-279.

Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics* 110(3): 681-712.

Mayne, J. (2001) "Assessing attribution through contribution analysis: Using performance measures sensibly." *The Canadian Journal of Program Evaluation* 16 (1): 1-2.

Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200-249.

Organization for Economic Cooperation and Development,; Wood, B; Kabel, D; Sagasti, F; Muwanga, N. *Synthesis Report on the First Phase of the Evaluation of the Implementation of the Paris Declaration*, Copenhagen, July 2008. [http://www.oecd.org/document/18/0,3343,en\\_2649\\_3236398\\_35401554\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/18/0,3343,en_2649_3236398_35401554_1_1_1_1,00.html)

Pratham Organization. "Annual Status of Education Report." (Pratham Resource Center:Mumbai), 2005.  
Dyck, Alexander, David Moss, and Luigi Zingales. 2008. "Media versus Special Interests." NBER Working Paper No. 14360.

Picciotto, R. (2007) "The New Environment for Development Evaluation," *American Journal of Evaluation*, 28 , (4), 509-521.

Ravallion, Martin. (2008) *Evaluation in the Practice of Development*. Policy Research Working Paper, Washington, DC: World Bank.

Shapiro, I. (2005) "Theories of Change". *Beyond Intractability*. Ed. G. Burgess and H. Burgess. Conflict Research Consortium, University of Colorado, Boulder, Colorado. <http://www.beyondintractability.org/essays/theories>

The Commission on the Measurement of Economic Performance and Social Progress (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. [http://www.stiglitz-sen-fitoussi.fr/documents/rapport\\_anglais.pdf](http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf)

Thomas, Vinod (2008). *"How Evaluation Must Help Improve Development Results."*