

Evaluations and Aid Effectiveness

by Pierre Jacquet¹

All bilateral and multilateral development institutions devote important resources to evaluating their operations. This activity is generally assigned to a dedicated unit placed outside the purview of the operational sphere so as to protect it from any vested interests; it is also often conducted in partnership with independent experts. Although they account for only a tiny proportion of public budgets, bilateral official development aid institutions probably even stand several steps ahead of many public administrations in thus contributing to the evaluation of public policies.² This brief paper, based on the experience of a bilateral development agency, claims that evaluations in fact fulfill several distinct, albeit related, valuable functions and that it is worthwhile to address each of them for their own sake as methods and organization need to be tailored accordingly. These are layman observations, however, that do not aim at constituting a “theory” of evaluations, but simply at providing some perspectives about an old but increasingly important function or set of functions in development institutions.

This paper does not primarily address the working of the World Bank Independent Evaluations Group (IEG, formerly OED—Operations Evaluation Department). I believe, however, that bilateral and multilateral development institutions all face similar challenges and should exchange more on their evaluation approaches and practices in order to improve on these practices and set common standards through a careful benchmarking process. Indeed, the description of IEG’s evaluation tools and approaches³ cover much of what I describe below as full-fledged “evaluation”, even though I propose here

Pierre Jacquet, Agence Française de Développement

Pierre Jacquet has been executive director (in charge of strategy) and chief economist at Agence Française de Développement (the French Development Agency) since 2002. He was formerly deputy director of the French Institute of International Relations in Paris and chief editor of its quarterly review *Politique Etrangère*. He is professor of international economics and chairman of the Department of Economics and Social Sciences at *Ecole nationale des ponts et chaussées*. Between 1997 and 2006 he was a member of the *Conseil d’Analyse Economique*, an independent advisory panel created by the French Prime Minister.

a much more explicit separation between the various functions that I identify. In analyzing these functions, I also try to highlight some organizational consequences. Not everything needs to be “independent” or conducted outside the operational sphere. And, insofar as scientific impact evaluation is concerned, there is a powerful case for some form of coordination among development institutions.⁴

1. Current evaluations: a multi-purpose activity

While there is a welcome attention currently given to the development of scientific impact evaluations geared to seriously assessing the actual impact of ODA financed operations on development, two related observations stand out from the experience of development institutions: first, not much has been done in terms of conducting genuine scientific development impact measurements so far; second, this is in fact not what most development institutions call “evaluation” in the first place. Yet, existing evaluation units play important roles that it is worthwhile to recognize and strengthen. They typically simultaneously fulfill three useful, different, functions, while admittedly not serving other, useful purposes that do belong to evaluation and that are in short supply in the international community of donors.

Building knowledge on processes and institutions

The first role is a cognitive role of building operational knowledge of processes and practices within a development institution and about how they interact with practices, behaviors and institutions on the receiving side. Additionally, evaluations build knowledge about the countries in which donors operate, their societies, their institutions, their needs, the quality of their governance. In short, they contribute to an empirical field knowledge that helps staff build field experience and become seasoned to the intricacies and complexity of development aid. This aspect of “evaluations” is important for the institution and needs to be conducted in-house because it is part of a necessary process of self-knowledge and self-education. It is also key to improving practices. Critiques comparing that process with scientific impact evaluation are based on a confusion of roles. What such evaluations do amounts to capitalizing on the existing experience

and creating transmissible knowledge from experience. They inform on processes and on the interaction between donor processes, practices and behaviors and institutions on the receiving side. They also point to shortcomings that can be remedied in subsequent operations. They can also be complementary to impact evaluations to the extent that they inform on why observed results may have been achieved.

A key aspect of evaluations, insofar as this first role is concerned, relates to the feedback process through which the results from these evaluations, however conducted, will inspire new actions. Substantial improvement needs to take place on feedback and retroaction, even though these dimensions have been recognized for a long time and have often led to substantial quality improvement. How to make feedback more systematic should be a top management priority. There is a natural tendency to focus on the “production of knowledge” aspect of evaluations rather than on how to use that knowledge in daily operations. Insufficient time and resources are devoted to that process and the structure of incentives is more favorable to conducting evaluations than to learning from them in the operational process. Indeed, a key to “results-based management” should not be as much about designing success, a rather elusive goal, as about learning from failures. This is certainly an area in which much progress needs to be achieved. A former United Nations Food and Agriculture Organization (FAO) officer for example described in his book⁵ how it took ten years for the FAO to abandon its recommendation that farmers use centralized storage facilities and villages establish cereal banks despite a record from evaluations that such centralization was inefficient and not viable.

While it is important, as we argue below, to add other dimensions to evaluations, it is also crucial to devote more time and resources to the feedback loop through which knowledge helps improve overall quality. This does not apply only to knowledge built through evaluations. An important aspect of quality management hinges on the ability of any institution to use all relevant knowledge in shaping its actions. Good practices must therefore ensure that decision processes do take into account results from past evaluations, but also current knowledge existing outside the institution and that needs to be identified and collected. Informing decision processes through

these various channels is a priority.⁶ A major difficulty, however, hinges on what is, or is not, relevant when taking a decision on whether to finance an operation, given the differences between past and current operations in terms of overall political, institutional, economic, social and technical environment.

Monitoring quality

Evaluations also fulfill an important monitoring role, that can be part and parcel of the practice of evaluations described in the preceding section, but that is of a different conceptual nature. It amounts to checking observed results against ex-ante expected ones and to monitoring the execution of operations: What was the objective of the program or project, was the money used for what it was supposed to, did it reach the intended beneficiaries, what was the time schedule for disbursements, how to explain any difference between actual versus expected results, and so on. This is an exercise in conformity and based on identifying good operational practices and on monitoring these practices. While the cognitive role alluded to above is best achieved through retrospective evaluations, quality needs to be monitored during execution and up to the end of any given operation.

As such, monitoring needs to be conducted within the operational sphere because this is where the relevant information is to be collected. Regular auditing must check that appropriate good practices and procedures have been followed. Monitoring also supplies useful information for the kind of evaluations that was mentioned in the preceding section. Here also, feedback, as a way to improve management, is an important dimension. It is central to quality management and control. Monitoring is thus an important instrument for top management.

External Accountability

As official development assistance deals with taxpayers' money, organizing accountability is of course a central task. There is, however, a necessary distinction to make between accountability and judgment. Development institutions are liable to provide an assessment of what they do and to communicate it to the outside. Accountability should amount to making all information available to the outside and let outsiders judge from that information and from their own perspective whether institutions make a good use of public money or not.

This aspect of evaluation systems, however, is open to challenge. There is an inherent conflict between the internal and external uses of information produced by evaluations. While the institution itself needs to learn from its own mistakes, it has little incentive to communicate outside on all potential mistakes even if it may gain in credibility in being quite open and transparent. This is why external evaluations are needed both to build institutional credibility and to form a judgment on the quality of any institution's operations. They should be commissioned and conducted in a fully independent way, outside institutions (or at least organized by them in a mutual, cooperative way so that peer pressure and peer review allow to counter the inevitable bias of having an institution judge its own actions). For example, the operations of the French Global Environmental Fund (FGEF), a Fund set up by the French government in parallel with the French participation in the Global Environment Facility (GEF) are thoroughly investigated before the decision to replenish the Fund, every three years. A team of independent auditors hired by the Board reviews the accounts, strategies and operations of FGEF.

There is a further dimension of accountability and judgment that could usefully be developed and that the current evaluation systems have not taken on so far, namely benchmarking of donors, multilateral as well as bilateral. Such benchmarking would be useful for two different sets of reasons. First, judgment of donor operations lack appropriate benchmarks. Quality of overseas development aid is better judged through comparison than in absolute terms. Benchmarking will better help inform taxpayers on whether their money was well used and whether it served useful purposes from their perspective. Second, benchmarking allows to compare the relative performance of donors and creates powerful incentives to improve efficiency in a world in which there is increasing competition among donors to collect development money, from a "market for aid" point of view, not only between public donors, but also between private foundations, NGOs, municipalities, etc.

May be even more than quality competition between donors, however, information on which donor does what best is also very valuable, as it may lead, within certain limits set by the global objectives of bilateral and multilateral assistance, to a possible natural division of

labor between donors. There are already some interesting initiatives in that respect. The Consultative Group to Assist the Poor (CGAP), a consortium of 31 public and private development agencies that work toward expanding access to financial services for the poor in developing countries through micro-finance projects, took a very interesting aid effectiveness initiative in 2002 in organizing a peer-review of its member activities. Results were discussed in several meetings and identified donors that performed well and others that did not. Interestingly, at least one donor found to perform poorly decided to retire from microfinance operations as a result of this peer review. Such an approach obviously could be replicated in other areas.

2. Scientific impact evaluations

However well done and professional, current evaluations are not up to the task of measuring the actual impact of development projects and programs. Impact evaluation is a demanding task, because it requires both careful observation of direct and indirect results and careful assessment of attribution of these results to the operation that is under-evaluation. Control groups are needed, and the whole process of impact evaluation should be thought of very early on, as early as the operation to be evaluated is itself identified. Recent progress in scientific impact evaluation methods (be they random assignment methods or other rigorous ones) now make it possible to learn useful lessons about the actual development impact of some operations, and there is a welcome move toward developing this kind of approach. Impact evaluations are surely not a panacea, but they give a renewed dimension to evaluations, more akin to applied research than to the ex-post assessment of operations traditionally undertaken in evaluation units. It fulfills a fundamentally different role of building scientific knowledge, and it is worthwhile to encourage its development.

Decades of development assistance have brought home the fact that development is a complex and poorly understood process. This is a powerful reason for focusing on empirical approaches that will highlight the kind of actions and policies that achieve results and those who do not and why. In turn, such knowledge is necessary to allow for more effective selectivity within development institutions, allowing them to focus on what

works and to avoid spending money on what does not. More than an evaluation of development institutions themselves, impact evaluation is a contribution to the provision of a global public good, namely knowledge on the development process.⁷

This public good character, together with the high cost involved in conducting scientific impact evaluations suggest that they will typically be under-supplied unless the coordination problem of who does what is solved. There is a further value added in coordination: better knowledge about what works in terms of development will have to be based not on empirical results from a single operation, but from comparing results from a number of operations in the same sector or with similar objectives. The way forward is setting up a system of cooperation between universities and development assistance donors toward joint impact evaluation, in which many donors take part, each of them contributing and all sharing the knowledge that is produced.

Available scientific methods, such as random evaluations, however, will work on some operations and much less on others. For cost as well as practicality reasons, not everything can undergo a scientific impact evaluation. Not everything is amenable to, say, random assignment like experimentation with drug use. For the very credibility of any exercise in scientific impact evaluation, therefore, it is crucial not to present it as a new religion, but rather as a contribution to better knowledge. For a start, what donors can usefully do is to initiate the process by selecting a few appropriate projects on which it is possible to build an operational cooperation between academic specialists and operational project officers, and by adopting a forthcoming, pragmatic approach. A number of initiatives deserve to be encouraged, notably the DIME project launched by the World Bank that proceeds along such lines.⁸

As argued above, however, when discussing the use of relevant knowledge in decision making, there will always be a need to keep a critical eye on the results from such scientific evaluations. It is unlikely that any scientific method of evaluation will allow to grasp all relevant factors in the interpretation of how a project or program fares in a given context, especially the institutional, human and societal dimension. What the method will help provide

is a scientifically informed knowledge base on actions, not a book of recipes about development.

Moreover, it is useful to keep in mind that development is more about processes than results. An objective of “impact” evaluation should thus be to help measure incremental improvement rather than final impact. Even when positive impact is reached and documented, the question remains about sustainability. From this point of view, there is a continuum of concerns between process and impact evaluations.⁹

3. Concluding Remarks

Part of the difficulty in debating the evaluation function in donor institutions is that a number of different tasks are implicitly simultaneously assigned to evaluation: building knowledge on processes and situations in receiving countries, promoting and monitoring quality, informing judgment on performance, and, increasingly, measuring actual impacts. Agencies still need their own evaluation teams, as important knowledge providers from their own perspective and as contributors to quality management. But these teams provide little insight into our actual impacts, and, although crucial, their contribution to knowledge essentially focuses on a better understanding of operational constraints and local institutional and social contexts. All these dimensions of evaluations are complementary. For effectiveness and efficiency reasons, they should be carefully identified and organized separately: some need to be conducted in house, some outside in a cooperative, peer review or independent manner. In short, evaluation units are supposed to kill all these birds with one stone, while all of them deserve specific approaches and methods.

There is a need to substantially buttress scientific impact evaluations, because they clearly exhibit public goods characteristics in terms of providing empirical knowledge on development. They require increased cooperation among donors and joint action. A number of initiatives have been launched recently, notably under the aegis of the World Bank. The focus on developing impact evaluations, however, should not obfuscate the need to considerably improve the operational feedback from evaluation results and more broadly from all available, relevant knowledge to operations.

Knowledge is not a scientific good only. Careful impact evaluation is a complement, not a substitute, to the non scientific, empirical approach that is also part of knowledge building and quality control. As for judgment of performance, this is clearly not a mission for the donor agencies themselves: Their responsibility is to be accountable, namely to provide all available, accurate and unbiased information on their operations. Assessment of performance needs to be totally externalized and should not be even undertaken under a contract directly commissioned by the donors themselves, except through a carefully organized peer review system.

As a multilateral institution with a clear commitment toward improving aid effectiveness and researching on development processes, the World Bank should contribute placing the role and format of evaluations higher on donors' agendas. It has taken a leading role in developing impact evaluations and engaging other donors in co-organizing a publicly available knowledge base about the results from such evaluations. In a recent report presented to Paul Wolfowitz, President of the World Bank, in the spring of 2005, Birdsall and Kapur¹⁰ notably recommend that the Bank should take the lead on independent evaluation of all aid spending. It should first, however, be also exemplary in taking part in existing peer reviews of donor operations in specific areas. Regretfully, the Bank did not participate in the CGAP peer review discussed above. Much is to be said in favor of a joint action by donors, alongside the CGAP example, to organize a rating of their operations in specific areas. As CGAP has demonstrated, it can provide very useful insights about what works and what does not and help donors become more selective in their operations.

Finally, development finance is not as much about picking out operations that work as about taking informed risks to discovering what works. Evaluations, along all the dimensions discussed above, are most useful to inform risk taking and decision making, rather than to act as substitutes to risk. The central, operational question is not whether operations similar to the one under consideration have been demonstrated to work in the past. In a dynamic, innovative, approach, it is rather whether all relevant knowledge has been called and taken into account so that the risk of development finance is carefully assessed.

Notes

1. I thank Pierre Forestier, Sebastian Linnemayr, Thomas Melonio, Jean-David Naudet and participants in the Center for Global Development symposium for useful comments, questions and suggestions. The usual caveat applies.

2. In particular, it is puzzling—and of questionable legitimacy—that we are seemingly putting more time and energy in trying to properly organize the evaluation of development projects than we seem to be in trying to assess the effectiveness of public policies in our own countries.

3. See www.worldbank.org/oed/.

4. For proposals along these lines, see William B. Savedoff, Ruth Levine and Nancy Birdsall, “When Will We Ever Learn? Recommendations to Improve Social Development through Enhanced Impact Evaluation,” Consultation Draft, Center for Global Development, Washington D.C., September 15, 2005; and also Levine and Savedoff in this volume. Through its Development Impact Evaluation (DIME) project, the World Bank has also undertaken an exercise about encouraging scientific impact evaluations both in the Bank and in partner development institutions, with the aim of collecting and sharing results so as to improve knowledge on several aspects of development processes.

5. Eberhardt Reusse, *The Ills of Aid. An Analysis of Third-World Development Policies*, (Chicago: University of Chicago Press, 2002).

6. This is a problem akin to the one studied in the pioneering work by Richard Neustad and Ernest May, *Thinking in Time: The Uses of History for Decision Makers* (New York: The Free Press, 1986).

7. See for example Esther Duflo, “Evaluating the Impact of Development Aid Programmes: The Role of Randomised Evaluations,” in *Development Aid: Why and How? Towards strategies for effectiveness*, Proceedings of the AFD-EUDN 2004 Conference, Notes and Documents No. 22 (Paris: French Development Agency, 2004).

8. The French Development Agency (AFD) has also decided to invest in scientific impact evaluation, starting with two projects, in microfinance and in health.

9. See the discussion in Smutylo, Terry (2001), “Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs”, Draft Learning Methodology Paper, Block Island Workshop on Across Portfolio Learning, 22–24 May 2001, Ottawa: International development Research Centre.

10. Nancy Birdsall and Devesh Kapur, co-chairs, *The Hardest Job in the World. Five Crucial Tasks for the New President of the World Bank*, in this volume.