



Current Challenges in Development Evaluation

SEAN CONLIN AND RODERICK L. STIRRAT
University of Sussex, UK

Over the past 20 years, the development industry has changed rapidly and now presents evaluators with a series of new challenges. This article considers this process of change. It argues that there has been a move away from a 'project-based' approach to a broader vision of development centred on the Millennium Development Goals, which has led to a change of the evaluation approach: from an assessment of outputs to an evaluation of results. Furthermore, the increasing complexity of ways in which the development process is conceptualized demands increasingly sophisticated and interpretive approaches to evaluation. Whilst there will remain space for positivist approaches in certain sectors of the industry, it is increasingly likely that other approaches will become more dominant.

KEY WORDS: development aid; evaluation approaches; evaluation challenges; narratives; policy outcomes

Introduction

Over the past 15 years, there has been a series of major shifts in development thinking and practice. This has involved new ways in which assistance is directed towards the developing world and new ways of conceptualizing the relationship between donors and recipients. Not surprisingly, these changes have created major challenges to those involved in evaluating development assistance.

In October 2006, a workshop was held in the University of Sussex to examine these changes. This drew together participants from a range of countries and institutions who are, in one way or another, involved in development evaluations.¹ Immediately after this workshop a smaller number of participants went on to deliver presentations in a session of the joint UK Evaluation Society/European Evaluation Society conference in 2006. This article is based on the presentations given at these two events, and also draws on the authors' knowledge and experience of development evaluation. It seeks to bring together the major strands of the discussions, identify the major challenges which have to be faced, and considers the various responses which have so far emerged.

The Context: The Changing World of Development Assistance

Since the mid-1990s, the development world has experienced a series of major changes. While the traditional roles of evaluation, primarily accountability and lesson learning, are still important, these have become more challenging in the new world of development. Today, evaluators are faced with increasingly complex situations and the demands being made upon them by their sponsors are correspondingly more difficult.

At the risk of caricature, until the 1990s, the 'project model' dominated development thinking and practice and provided the context for the theories and methods of development evaluation.² Aid was generally delivered in the form of projects, a tightly bounded set of activities that typically took three years to complete. The stress in these projects was on project staff producing 'deliverables'. How these deliverables were to be delivered was set out in a 'logical framework' which defined the presumed links between the inputs, outputs and overall outcomes, as well as the assumptions underlying these links. Evaluation tended to focus on whether or not these 'deliverables' had been delivered and whether the assumptions had held. There was relatively little attention to 'impact assessment' which would have required a longer time frame and ranged wider than the tightly defined objectives of the project.³

This project model of development was associated with a particular political and institutional context of development assistance where donor interests had primacy. Donor agencies 'owned' the projects rather than the developing countries where they were located. Donors provided project funds to resolve a particular constraint on development but the developing country government was expected to provide for the ongoing maintenance and operation. Within this frame, donor priorities shifted from large infrastructure projects to rural development projects as international theories of development (or fashion) changed. Developing country institutions were considered weak so donors provided technical assistance, usually experts from the donor country. These experts undertook most of the project management; separate project management units were common. These management units often had their own monitoring and evaluation sections which concentrated on monitoring inputs and outputs, less attention being paid to the overall outcomes. These arrangements were understandable for a donor-owned time-bound project, after which the donor could hand over responsibility for ongoing activities to the developing country government. At best, donor projects were islands of excellence demonstrating what might be achieved; at worst, they were unsustainable and undermined local systems.

Since the 1990s there has been a major shift in the delivery of aid assistance away from donor designed and managed projects. Without going into the historical detail, this can be associated with the end of the Cold War, theoretical critiques of development from the right and left, globalization, increased importance of trade and private investment, aid fatigue among donors and structural adjustment.⁴

Overall, the emerging architecture of aid involved a number of trends, and we focus on four of these. The first of these trends has been the rise of 'management for results'. Traditionally, much of development assistance took the form of project

assistance where the primary stress was laid upon achieving outputs. From the early 1990s, there has been an increasing stress not on outputs but on impact and the effect of development assistance. Perhaps the most obvious manifestation of this shift has been the move from project-based logical frameworks into a 'results framework' for use in the new world of development interventions (DAC Development Partnership Forum Report, 2002; Van den Berg, 2005).

Second, there has been an increasing stress on targets culminating in the agreement on the 'Millennium Development Goals' (MDGs) at the Millennium Summit of 2000.⁵ Developed during the 1990s, this set of goals provides an overarching framework within which all development activity should fit. The stress in these goals is on objectives (or more precisely, a set of indicators which link to objectives) such as lowering the number of people living in extreme poverty, increasing literacy rates and decreasing child mortality. What is important about these goals is not just the stress on objectives rather than outputs, but that they are also associated with a movement away from project-orientated development interventions to a much wider view of what development assistance should do and how it should be delivered.

This is related to the increased stress on 'partnership', culminating in the 2005 'Paris Declaration on Aid Effectiveness'.⁶ This document codified already emerging principles for development partners. The first principle is the *ownership* by the developing country of its development strategy. Second, donors should *align* their activities behind the country's strategy. This implies that donors work towards *harmonizing* their different approaches and that greater attention should be given to *management for results*. The final principle is *mutual accountability* for achieving the results.

These changes in the structure of the aid community have created a new political and institutional context for development evaluation and pose new methodological challenges for the evaluator. Here we focus on four of these challenges:

- managing for results and the Millennium Development Goals;
- partnership principles;
- new aid modalities such as sector approaches and general budget support;
- going beyond aid to policy coherence.

As far as 'management for results' is concerned, the evaluator now has to move far beyond considering whether or not the 'deliverables' have been achieved to consider whether or not 'deliverables' have had the impact that was hoped for them. This obviously raises major problems in attribution: how far can one be certain that particular changes have been the result of particular development interventions? The influence of other factors outside the control of development policy-makers and practitioners becomes increasingly important and difficult to determine. Not only is this a methodological challenge to evaluators, but it is also a political challenge to donor agencies concerned to show taxpayers that aid funds are being used effectively. These challenges are increased by the use of the MDGs to define desirable impacts and the rise of a much wider view of what development assistance should do and can do (OECD, 2002).

The second trend complicating the role of the evaluator is the changing relationship between aid givers and aid receivers, and a new emphasis on partnership. Previously, evaluations were planned, implemented and produced for donors, but increasingly the evaluation process is seen as involving all the partners involved. Furthermore, what is being required of evaluators by these various partners is increasingly complex, and evaluators can now find themselves working to a range of audiences and masters. Older principles of accountability and conditionality are now replaced with an increasing stress on the learning functions of the evaluation process. Evaluations are increasingly viewed as contributing to 'empowerment', which is now a central theme in contemporary development thinking. And in practice, while evaluators have always worked in a political context, the demands being made upon them are increasing in variety and significance.

One of the manifestations of this change in emphasis has been an increased interest in 'donor alignment and harmonization'. If the overall objectives of development (as enshrined in the MDGs) are shared by all partners, it becomes incumbent on donors to harmonize their activities and align them with the partner country's own development strategies to ensure aid effectiveness. While donor agencies are trying to harmonize their practices, the problem facing evaluators is how to harmonize the often contrasting systems of evaluation employed by donors and partner countries, a task being undertaken by the DAC network on Development Evaluation.

The third trend, associated with this emphasis on alignment and harmonization, has been a change in the modalities of aid delivery, with increased stress on 'sector wide approaches' (SWAPs) and a move away from projects. Here, the focus of aid delivery is on assisting the sector as a whole to achieve overall objectives, usually phrased in terms of the MDGs, with donors and government identifying the kinds of policy reform and other activities required. The sector approach addresses a wide range of activities – from reforming the regulations in the sector to improving physical infrastructure and supporting training and capacity development. Although there has been a transitional phase in many countries where separate projects are aligned with sector objectives, increasingly donor assistance takes the form of pooled support of both financial and technical assistance. Thus the role of individual donors becomes much less clear and the evaluator's task much more complex. But this is not the only problem facing today's evaluators. The shift to SWAPs has only exacerbated the problem of attribution in that it is increasingly difficult to disentangle the results of donor assistance from the overall processes at work in any particular sector.

Moving beyond sector support, there has been a move among some donors to the much wider field of 'general budget support'. Here, donor assistance is directed towards supporting the general role and activities of governments in developing countries. While there is widespread acceptance among donors that SWAPs are more effective than a series of separate projects, the claim that general budget support increases aid effectiveness is not fully accepted by the donor community. This is due to a number of reasons, mainly associated with risk. As far as the evaluator is concerned, the challenge of reaching firm conclusions as to attribution of results to inputs and the chain of causation becomes more and more difficult.

Fourth, and finally, the increasing recognition of the importance of trade, investment and other economic and social activities in the development process has led to a growing interest in broader issues of development effectiveness. The argument is that development assistance is only one of many ways in which the developed and the underdeveloped world relate to one another, and that trade and private investment are in quantifiable terms much more important strands in this relationship. In similar fashion, development researchers have drawn attention to the importance for long-term development of factors such as security, migration, and climate change (OECD, 2005). Thus any realistic attempt to evaluate the effectiveness of aid has to be seen in the wider context of development effectiveness. Increasingly evaluators are being asked to assess how the wider policies of trade, finance and other areas intersect with the world of aid to generate particular impacts on the development process.

Overall, the situation facing development evaluators today is vastly more complex than it was 15 or even 10 years ago. Shifts in the nature of aid, coupled with changing relationships between donors and recipients, have made it much more difficult to produce neat and firmly demarcated forms of evaluation. Forms of evaluation based on models derived from project logical frameworks are no longer appropriate to the fluid and complex forms of development assistance in the 21st century, nor can they deal with the increasingly wide range of stakeholders involved in the development process. Not surprisingly, this has led to talk of an 'evaluation gap' in some quarters and the need for greater inputs into development evaluation (Center for Global Development, 2006).

Meeting the Challenge: Methodological Issues

There is, of course, still a demand for what might be called 'traditional' forms of evaluation. Thus organizations such as the MIT Poverty Action Laboratory and the working group of the Centre for Global Development see 'randomized scientific experimental trials' as the gold standard for impact evaluations and the standard which all should aspire to. Yet this form of evaluation is relatively rare in the development world (Rugh, 2006). This is in part a matter of expense, but there are other problems. They are only appropriate in a small minority of situations where there are no practical or ethical impediments to randomization and where there are no ethical problems in excluding a control group from the development intervention. More importantly, although the use of such techniques works well in relatively simple contexts (e.g. the impact of drugs on health; the impact on literacy of increased educational expenditure), in the world of development today, the range and variety of factors which influence development trajectories are simply too complex to be dealt with in terms of simple cause and effect models. Where there is perhaps space for such methods is in approaches to development which stress 'development projects'. And thus it is not surprising that the greatest faith in such methods is in the United States where development projects continue to dominate the development landscape.

Similarly, there is still a high degree of commitment to 'theory-based evaluation', but here the problem is that only a minority of development interventions can be

simply dealt with through recourse to the original (or even revised or modified) project plans. In the complex world of SWAPs, donor harmonization and budgetary support, there is less room for the effective use of logical frameworks or other project planning tools and a need for more flexible and less constrictive methods. Most of these in one way or another build upon past experience but at the same time explore new and innovative pathways.

One of the main challenges mentioned is the problem of how to organize and implement evaluation in complex situations involving a range of partners. Here, a number of new approaches can be identified.

The first of these is the growing popularity of joint evaluations involving a number of different agencies. These began as 'synthesis studies' which brought together the results of evaluations by different donors. More recently they have evolved into joint evaluations at a sector- or country-level involving a range of donors and make no attempt to isolate the effect of each donor's activities (Lithman, 2006). While having clear advantages over separate evaluations by different donor organizations, there are major problems both logistically and in terms of harmonizing evaluation practices between the various agencies involved.

The second derives from an academic tradition of social network analysis and is particularly suited to situations where a range of organizations, often of very different types, are all involved in a common endeavour but often with very different objectives. The aim of evaluations based on network models is to analyse how decisions are made in such situations where there is no clear 'command and control' managerial structure and to suggest to the various organizations involved how common objectives might be arrived at or alternatively how complementary objectives might be established. Here, the primary aim of the evaluation is not accountability but learning in order to increase future effectiveness (Davies, 2006).

Another approach, also based on the idea of social networks, is 'outcome mapping'. This method focuses on how information and knowledge passes through networks and whether or not the original inputs have any developmental impact. Again, this method is particularly suited to situations where a range of partner organizations are involved and it seeks to identify barriers to dissemination as well as determining the factors which inhibit or encourage the successful paths of influence – in both directions (Davidson, 2000; Earl, 2006; Earl et al., 2001; Stirrat, 2003).

These various approaches to development evaluation build upon older approaches which were current when projects were the dominant form of development aid. They work in terms of how a specific input or activity reverberates at a social and institutional level. Although there is an attempt to involve all partners in the evaluation endeavour, in practice they tend to be based on the needs and interests of the donors. Clearly there is an interest in the development community to move beyond this position. In particular, the movement away from projects as the dominant form of aid delivery has led to a great deal of interest in how evaluations might be designed to deal with the complexities of SWAPs and budgetary support programmes and at the same time meet the demands of a widening and much more varied audience for the products of evaluations.

Such issues are clearly visible in the context of poverty reduction interventions. These frequently involve whole agencies or groups of agencies, and the traditional 'independence' of the evaluators from the implementers becomes blurred. The complexity and processual nature of such interventions make it difficult to pre-plan and integrate data collection into the management process while counterfactuals are not viable approaches. The overall result is that traditional evaluation approaches are extremely difficult to operationalize and that what is produced often lacks conviction.

One way in which this issue of complexity has been addressed (or perhaps avoided) is through the evaluation of agencies. This includes the DAC Peer Review process as well as a series of evaluations of particular international financial institutions such as IDA, the Asian Development Fund and the African Development Fund. These, in the main, work through assessing institutional performance against a series of 'best practice' standards, themselves the result of previous evaluations (African Development Bank, 2004; Liverani and Lundgren, 2007; World Bank, 2002).

An alternative approach is through the development of 'country programme evaluations' which focus on the overall aid structure within a particular partner country. This approach, championed by various agencies such as the UN family and the Development Assistance Committee (DAC) of the OECD, has certain strengths in that it is often based on a participatory and rights-based approach (Back, 2006; Freeman, 2005). But there are problems in that so far it has proved difficult to shift the focus of evaluation from a stress on 'outputs' to the strategic and programme-setting levels.

A recent attempt to evaluate Uganda's Poverty Reduction Strategy adopts a different strategy (Kakande, 2006). One of the core problems addressed in this evaluation was the multidimensional nature of poverty. The evaluation focused on a group of 31 indicators of output and impact as well as another set of indicators to assess inputs and value for money. While the factors which determined the choice of indicators is not clear, and there were problems in aggregating data, the result of the exercise significantly increased the Ugandan government's ability to manage the poverty reduction programme. Even so, there remain major problems in evaluating PRSPs (Holvoet and Renard, 2007).

A more ambitious approach, developed in the context of evaluating general budget support across a range of countries, has been the 'Enhanced Evaluation Framework'. This seeks to establish a broad framework which sought to integrate in the analysis a wide range of factors. Thus, while using a theory-based approach, it also considers policy effects and the flows of funds and institutional factors, as well as assessing the role of parallel and complementary inputs. But despite the sophistication of the approach, the development of a 'causality map' and a range of robust conclusions, there are still major problems. Attribution remains a major issue, as do issues concerning how one can evaluate an ongoing approach as distinct from a strategy, programme or project. Perhaps the most difficult question is determining where the boundaries of the analysis should be drawn, given the complex intermingling of financial and technical inputs within a changing policy framework (Jacobs, 2005; Lister, 2006).

This raises the more general issue of what should be the subject of evaluation in the development context. One of the recurring problems is how far evaluators should restrict their activities to the activities and impacts of development agencies. An obvious problem is that it is often difficult to distinguish the impacts of these forms of development interventions from the wider processes of which they are only a part. Furthermore, the general adoption of the MDGs (in particular MDG 8, which stresses the importance for development of policy coherence) has led to a much greater stress on the development process as a whole and the achievement of a wide range of developmental targets. But development assistance in itself is only a minor component in the range of flows between the developed and the developing world and is dwarfed by trade, by foreign direct investment and by remittances. Thus it can be argued that the focus of development evaluation should not be restricted to development assistance but to this much wider set of forces. The evaluation industry thus has to develop a new set of priorities centred on policy coherence and how the different strands of policy in the developing world covering not just aid but also trade, investment and migration, as well as the nature of global public goods (e.g. environmental factors) impinge on the development process. Furthermore, renewed attention should be paid to evaluation at the country level, once again focusing on the whole range of factors which impinge on a country's development (Picciotto, 2005, 2006).

Clearly, this will involve a major rethinking in the nature of development evaluation. Leaving aside for one moment the institutional challenges which such an approach to evaluation involves, there is also a host of methodological and theoretical issues which will have to be addressed before 'development' evaluation rather than 'aid' evaluation becomes viable. Elements of what would be required are clearly visible: for instance, the importance of theory-based evaluation, of objectives-based approaches, of meta-evaluation techniques and the use of case studies. But the sheer scale of the endeavour and the problematic nature of producing convincing attributional arguments make development evaluation the major challenge of the 21st century.

In sum, once upon a time the role of the evaluation specialist was relatively simple. As long as development was defined in terms of aid and aid in terms of projects, the evaluator was able to focus on relatively well demarcated and constructed sets of factors. But as understandings of development have become more sophisticated and broad-ranging, so the task has become increasingly complex. How far the existing range of techniques employed by evaluators is suitable for the new tasks they face is an open question.

This in effect brings us back to the beginning of this section: is there a continuing role for the 'scientific' model of evaluation epitomized by randomized trials? Given not just the increasing complexity of the development world but also the increasing range of objectives and criteria employed in that world, does a positivist approach make sense? Perhaps the time has come to recognize that interpretive approaches which owe more to history than to experimental science are better suited to the world of development (Van der Berg, 2005). Perhaps the time has come to recognize the importance of narratives which make sense of what is

observed, narratives which are never final products but always, like development, in a state of 'becoming' (Dart and Davies, 2003; Hulme, 1997).

Meeting the Challenge: Institutional Factors

The shifts in the nature of the development industry have led to new stakeholders and new interest groups becoming increasingly interested in the processes of evaluation. Yet precisely who these are, what sorts of interest they have in evaluation, the character of the relationship between them and whether or not they have the capacity for evaluation, all remain open questions.

In the past, the dominant market for the products of the evaluation industry consisted of the donors. The twin objectives of accountability and learning dominated evaluations, and donors determined almost all aspects of evaluations from choice of evaluators through methods used to the form and distribution of the final products, the evaluation reports. The demands of the market were clear: to make sure that resources had been used in a 'reasonable' fashion, and that 'lessons' could be learnt to make future development interventions more effective. How far lessons can actually be learnt is, however, a matter of some debate (Cracknell, 2001; Marra, 2000; Picciotto, 2002; Schaumberg-Muller, 2005). In effect, evaluation was 'supply-driven' by the donor agencies and, as far as people in developing countries were concerned, something which was imposed from outside.

The situation today is somewhat different. With the movement towards 'partnership' between developed and developing countries, and with political and social changes in those countries, there is an increasingly wide set of potential 'consumers' for the products of the evaluators (Gariba, 2006). Of course, there is still a demand from the donor countries for accountability, a demand which seems to be growing as the aid shifts from more visible 'projects' to less visible 'programmes'. But demand is also growing in recipient countries, not only from governments and government departments, but also from civil society and the NGO communities in developing countries. This creates a new and complex landscape within which the evaluator has to work and an increasingly complex set of demands which have to be met. So, for instance, does meeting these widening sets of interests necessitate involving an ever-widening range of representatives in the evaluation team? How can the increased transaction costs which partnership involves be limited? How can evaluators produce products which meet the differing demands of, say, donor governments on the one hand and NGOs, civil society organizations and state institutions in partner countries on the other? How far does the call for 'shared objectives' coupled with 'distinct accountabilities' allow a 'partnership' approach to evaluation – or does it remain little more than attractive rhetoric?

This inevitably involves evaluation taking on an increasingly visible political profile. Even though evaluations have always been 'political' it was in the past possible to maintain a facade of 'neutrality' and 'objectivity', but as the demands from increasingly vociferous civil society organizations grow, so this stance becomes increasingly difficult to maintain. This is most obvious in recent discussions of 'empowering evaluations' where evaluation is seen as a process through which

marginal and disempowered groups are able to gain skills and influence through their involvement in evaluation (Fetterman and Wanderson, 2004; Fetterman et al., 1996; Holte-McKenzie et al., 2006; Schnoes et al., 2000). It is also apparent in the changing ways in which 'accountability' is defined. No more is it simply a matter of accountability to donors, but also of downward accountability to beneficiaries. Here, what evaluators do and don't do is increasingly a matter of political forces, but more importantly perhaps, the distinction between 'evaluation' and 'management' becomes increasingly blurred, and the degree to which evaluators can maintain a semblance of independence becomes increasingly problematic (Naidoo, 2006).

Among the donors, there are some signs that the prevalent structure which placed evaluation departments within donor organizations is under pressure. In Sweden, for instance, consideration is being made of the possibility of moving development evaluation out of the hands of SIDA, the major Swedish development organization, and making it one activity among several which are the responsibility of an independent evaluation organization. This would appear to be one way in which the evaluation process can start answering the needs of a wider range of stakeholders and not just the donor organization. An alternative would be to build up the capacities of independent development companies, analogous to independent audit companies, which again stand outside the institutional framework of the development industry, although the problems of who pays for the evaluations would remain significant.

One alternative is to increase the stress on enhancing evaluation capacity in the developing world, although it is often not clear as to whose capacity should be developed: development organizations, political groupings or the institutions of civil society. Efforts in this field involve a number of related themes. The first of these is to ensure that data are available on which evaluations can be based, and this in turn involves increasing the capacity of organizations in developing countries to generate accurate and timely data. The second is training, either through formal means or through broadening the membership of evaluation teams. Third, there is a recognized need for stakeholders outside of the government to have access to the skills and techniques of established evaluators (Ofir, 2006). Finally, there is the issue of 'participatory evaluation' which links to issues of 'empowerment' and stakeholder involvement (Gregory, 2000).

Clearly, any radical moves in the way in which evaluations are organized will further strengthen the need for some form of quality control in development evaluation. At present there are no generally agreed standards; nor are there means through which the quality of evaluators' work can be judged. This raises the question of how far 'evaluation' should be seen as a 'discipline' with its own academic standards, body of theoretical knowledge and its own particular methodological approaches, or whether it is better to see it as a 'profession', a body of people abiding by a code of conduct but not necessarily sharing a common theoretical outlook, relying on a range of technical specialisms for inputs into the overall evaluation. This appears to lie behind the increasing interest in the establishment of international evaluation organizations (Dimitrov, 2006; Love and Russon, 2000; Mertens, 2005). On one level this involves organizations such as IDEA but more ambitious proposals include the establishment of independent

evaluation institutions at an international level. What this would imply for the strengthening of evaluation capacities in developing countries or the involvement of non-government organizations remains unclear.

Evaluation and the Policy Process

In the end, evaluation is of little more than academic interest unless it feeds into policy-making and has a positive impact on the development process. It has to be said that, while there have been evaluations which have had a major impact (DANIDA, 2004; World Bank, 2005; World Commission on Dams, 2000), only too often in the past the messages generated by evaluators have had little influence on policy-makers either within agencies themselves or within the wider development world. Part of the problem has been a lack of timeliness – evaluations are often too late to be relevant. There has also been a lack of credibility of evaluators and evaluations amongst policy-makers, and a tendency by development organizations to treat evaluation as a necessary but somewhat marginal activity. A third problem has been failure in communications. The published evaluation documents of some agencies are remarkable only by their unexciting format and dismal standards of presentation.

One way in which development evaluators are attempting to overcome these problems is by widening stakeholder involvement in both commissioning and implementing evaluations. This can involve not only donors and partner-governments but also NGOs and various forms of civil society organizations. This generates a greater sense of ‘ownership’ among such bodies in the outputs of evaluation and also ensures that evaluations are directed towards issues which are of direct relevance to policy-makers. The problem of course is to maintain the independence of evaluators and prevent the ‘capture’ of the evaluation process by policy-makers and implementers.

A second method is to maintain as far as possible the independence of evaluators while placing evaluation more firmly within the development process. The traditional ‘post-project evaluation’ might have made sense in terms of a development world dominated by time-restricted projects, but the new world of development is an ongoing process. Evaluations have to recognize this change and evaluators have to see their activities and their outputs as feeding into this process rather than as being reflections on a completed process.

One of the issues which has to be addressed is the present push towards ‘evidence-based policies’. Here, the potential role of evaluation is clear: that it can provide the ‘evidence’ on which to base policies (Martin, 2005; Shadish et al., 2005; Stek, 2003). Yet, as we have argued earlier, it is not clear in the complex world of aid (let alone development) that evaluations can provide the sort of evidence that the proponents of evidence-based policies demand. The complexities of the situation are too great and the problems of attribution too complex for any satisfactory and persuasive set of causative relationships to be established. And as we have pointed out, the ‘gold standard’ of randomized comparisons is simply not ethically or practically possible in the majority of situations which the development evaluator has to address. This implies that theory-based evaluation, the

construction of narratives and the use of the historical method rather than positivist science, has to be made legitimate to policy-makers. One of the major challenges facing the development evaluation industry is thus of persuading policy-makers of the validity of the methods increasingly used in evaluation, methods which recognize the complexity of the contemporary situation.

But the greatest challenge is how evaluation can feed into the pursuit of the Millennium Development Goals. In general development evaluation still focuses on individual agencies or countries, either through evaluations of donor countries' aid efforts or through aid activities and their impacts in 'partner' countries. Only rarely does the gaze of the evaluator move beyond national or bilateral horizons, one example perhaps being the African Peer Review mechanism managed by NEPAD. Yet given that achieving the MDGs depends on a host of transnational factors and processes, the ultimate challenge of development evaluation is to influence policy-makers and decisions at the transnational level. Independence has to be maintained – but at the same time evaluators have to enter the political arena not just as neutral observers and reporters but as engaged activists.

Conclusion

It would be overdramatic to talk of a 'crisis' in development evaluation. But certainly there are major challenges which have to be addressed in the near future. As the overall architecture of aid and development has changed, so old methods of evaluation are no longer satisfactory. The result is the present series of lively debates and experiments as theorists, institutions and practitioners attempt to generate new methods and new organizational structures. In some ways it is clear that there will be no one solution. Different approaches will work in different contexts. But the overarching issue for evaluators is one of legitimacy: how can they persuade the various stakeholders who are now involved in the evaluation process of the legitimacy of these new methods? Older positivistic methods seem less and less able to deal with the contemporary situation. Yet can the 'interpretive' methods which are becoming more popular gain the legitimacy needed for evaluations to have any real effect? In a world which talks in terms of 'evidence-based policies' this will be a major challenge. While the Centre for Global Development has raised legitimate points about the problems facing development evaluation, the solutions they propose are outdated.

Notes

The authors wish to thank the Department of Anthropology and the School of Social Sciences and Cultural Studies at the University of Sussex, the UK Department for International Development, the Swedish International Development Agency and Irish Aid for their financial support both to the Sussex workshop and meeting the costs of overseas participants at the joint UKES/EES conference in 2006. They would also like to thank Dinah Rajak, who was responsible for the organization of workshop activities.

1. The presentations made at the Sussex workshop are available at <http://www.sussex.ac.uk/development/1-4-9-1-1.html> (consulted 7 November 2007).

2. For two recent treatments of the changing 'architecture of aid', see World Bank, 2007 and Burall et al., 2006. For an overview of evaluation in the World Bank, see Stek, 2003.
3. Many examples of donor evaluation studies by OECD member countries are available at the DAC Evaluation Resource Centre. http://www.oecd.org/document/63/0,2340,en_35038640_35039563_35067327_1_1_1_1,00.html (consulted 7 November 2007)
4. Donor policy documents frequently discuss the factors influencing change in the development context. See for instance DANIDA, 2005; DFID, 2000, 2006.
5. The Millennium Declaration can be found at <http://www.ohchr.org/english/law/millennium.htm> (consulted 7 November 2007). For a discussion of the implications of the MDGs for evaluation, see Kusek et al., 2005.
6. The OECD set out its vision of partnership in 1996: *Shaping the 21st Century: The Contribution of Development Co-operation*. In 2005 the DAC High Level Forum agreed the Paris Declaration on Aid Effectiveness which can be accessed at <http://www.oecd.org/dataoecd/11/41/34428351.pdf> (consulted 7 November 2007).

References

- African Development Bank (2004) *Stepping up to the Future: An Independent Evaluation of African Development Fund VII, VIII and IX*. Tunis: African Development Bank.
- Back, L. (2006) 'Challenges to Evaluation Capacity: Development through Country Level Evaluations', URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Burall, S. and S. Maxwell, with A. R. Menocal (2006) *Reforming the International Aid Architecture: Options and Ways Forward*. London: Overseas Development Institute Working Paper 278.
- Center for Global Development (2006) *When Will We Ever Learn? Improving Lives through Impact Evaluation. Report of the Evaluation Gap Working Group*, URL (consulted 7 November 2007): <http://www.cgdev.org/content/publications/detail/7973>
- Cracknell, B. E. (2001) 'The Role of Aid-Evaluation Feedback as an Input into the Learning Organization', *Evaluation* 7(1): 132–45.
- DAC Development Partnership Forum Report (2002) 'Managing and Measuring for Development Results and Aid Effectiveness, Paris, 11–12 December 2002', *DAC Journal on Development* 4(1): 1–5.
- DANIDA (2004) *Meta-Evaluation: Private and Business Sector Development Interventions*. Copenhagen: Ministry of Foreign Affairs.
- DANIDA (2005) *Globalisation: Progress through Partnership*. Copenhagen: Ministry of Foreign Affairs, URL (consulted 7 November 2007): <http://www.amg.um.dk/NR/rdonlyres/BD805561-B40B-462D-BC2F-C541823EFB29/0/GlobalisationProgressThroughPartnership.pdf>
- Dart, J. and R. Davies (2003) 'A Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique', *American Journal of Evaluation* 24(2): 137–55.
- Davidson, E. J. (2000) 'Ascertaining Causality in Theory-Based Evaluation', *New Directions for Evaluation* 87(1): 17–26.
- Davies, R. (2006) *Networks and Partnerships: Some Tools for Representing and Analysing Networks*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Dimitrov, T. (2006) *Networks and Partnerships: The Role of IDEAS, the International Development Evaluation Association*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- DFID (2000) *Eliminating World Poverty: Making Globalisation Work for the Poor*. London: DFID.

- DFID (2006) *Eliminating World Poverty: Making Governance Work for the Poor*. London: DFID.
- Earl, S. (2006) *Behaviour Changes as Development Outcomes. An Introduction to Outcome Mapping*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Earl, S., F. Carden and T. Smutylo (2001) *Outcome Mapping: Building Learning and Reflection into Development Programs*. International Development Research Centre (IDRC), URL (consulted 7 November 2007): http://www.idrc.ca/en/ev-9330-201-1-DO_TOPIC.html
- Fetterman, D. and A. Wandersman, eds (2004) *Empowerment Evaluation Principles in Practice*. New York: Guilford Press.
- Fetterman, D. M., S. Kaftarian and A. Wandersman, eds (1996) *Empowerment Evaluation: Knowledge and Tools for Self Assessment and Accountability*. Thousand Oaks, CA: SAGE.
- Freeman, T. (2005) *Country Programme Evaluation in an Era of Change*. Evaluation Working Papers. New York: UNICEF.
- Gariba, S. (2006) For Whom the Bell Tolls: Improving Accountability for Policy through Evaluation, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Grahn, R. (2000) 'A Study into Methods for Tracking the Research Uptake and Developmental Impact of NRSP Funded Research Projects', unpublished manuscript.
- Gregory, A. (2000) 'Problematizing Participation: A Critical Review of Approaches to Participation in Evaluation Theory', *Evaluation* 6(2): 179–99.
- Holte-McKenzie, M., S. Forde and S. Theobald (2006) 'Development of a Participatory Monitoring and Evaluation Strategy', *Evaluation and Program Planning* 29(4): 365–76.
- Holvoet, N. and R. Renard (2007) 'Monitoring and Evaluation under the PRSP: Solid Rock or Quicksand?', *Evaluation and Program Planning* 30(1): 66–81.
- Hulme, D. (1997) 'Impact Assessment Methodologies for Microfinance: A Review', paper prepared for a meeting of the CGIAP working group on Impact Assessment Methodologies, URL (consulted 7 November 2007): http://www.ids.ac.uk/impact/resources/introduction/Hulme_IA_meth_review.pdf
- Jacobs, C. (2005) 'Evaluating the Comprehensive Development Framework in Kyrgyz Republic, Central Asia: Magic Bullet or White Elephant?', *Evaluation* 11(4): 480–95.
- Kakande, M. (2006) 'Evaluating Poverty Reduction Strategies: Uganda's Experience', URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Kusek, J. Z., R. C. Rist and E. M. White (2005) 'How will we Know the Millennium Development Goal Results When We See Them?', *Evaluation* 11(1): 7–26.
- Lister, S. (2006) 'Evaluating General Budget Support', URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Lithman, E. (2006) *Trends in Development Evaluation: Changing Approaches*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Liverani, A. and H. E. Lundgren (2007) 'Evaluation Systems in Development Aid Agencies: An Analysis of DAC Peer Reviews 1996–2004', *Evaluation* 13(2): 241–56.
- Love, A. J. and C. Russon (2000) 'Building a Worldwide Evaluation Community: Past, Present, and Future', *Evaluation and Program Planning* 23(4): 449–59.
- Marra, M. (2000) 'How Much does Evaluation Matter? Some Examples of the Utilization of the Evaluation of the World Bank's Anti-Corruption Activities', *Evaluation* 6(1): 22–36.
- Martin, S. (2005) 'Evaluation, Inspection and the Improvement Agenda: Contrasting Fortunes in an Era of Evidence-Based Policy Making', *Evaluation* 11(4): 496–504.
- Mertens, D. M. (2005) 'The Inauguration of the International Organization for Cooperation in Evaluation', *American Journal of Evaluation* 26(1): 124–30.

- Naidoo, I. (2006) 'The Influence of the PSC on Policy and Governance in South Africa', URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- OECD (2002) *Action for a Shared Development Agenda*, URL (consulted 7 November 2007): http://www.oecd.org/document/46/0,3343,en_2649_33721_2088942_1_1_1,00.html
- OECD (2005) *Policy Coherence for Development: Promoting Institutional Good Practice*. Paris: OECD.
- Ofir, Z. (2006) *Developing Evaluation Capacity for National Poverty Reduction Strategies*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Picciotto, R. (2002) 'The Logic of Mainstreaming: A Development Evaluation Perspective', *Evaluation* 8(3): 322–39.
- Picciotto, R. (2005) 'The Evaluation of Policy Coherence for Development', *Evaluation* 11(3): 311–30.
- Picciotto, R. (2006) *The Global Dimension of Development Evaluation*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Rugh, J. (2006) *Addressing the Evaluation Gap*, URL (consulted 7 November 2007): <http://www.sussex.ac.uk/development/1-4-9-1-1.html>
- Schaumberg-Muller, H. (2005) 'Use of Aid Evaluation from an Organizational Perspective', *Evaluation* 11(2): 207–22.
- Schnoes, C. J., V. Murphy-Berman and J. M. Chambers (2000) 'Empowerment Evaluation Applied: Experiences, Analysis and Recommendations from a Case Study', *American Journal of Evaluation* 21(1): 53–64.
- Shadish, W. R., S. Chacon-Moscoso and J. Sanchez-Meca (2005) 'Evidence-Based Decision Making: Enhancing Systematic Reviews of Program Evaluation Results in Europe', *Evaluation* 11(1): 111–22.
- Stek, P. (2003) 'Evaluation at the World Bank and Implications for Bilateral Donors', *Evaluation* 9(4): 491–7.
- Stirrat, R. L. (2003) *Methodological Issues in Identifying the Impact of Research Projects: A Review of the Literature*. Stirling: Institute of Aquaculture, URL (consulted 7 November 2007): <http://www.dfid.stir.ac.uk/Afgrp/program/impact.pdf>
- Van den Berg, R. D. (2005) 'Results Evaluation and Impact Assessment in Development Co-operation', *Evaluation* 11(1): 27–36.
- World Bank (2002) *IDA's Partnership for Poverty Reduction: An Independent Evaluation of Fiscal Years 1994–2000*. Washington, DC: World Bank.
- World Bank (2005) *Influential Evaluations: Detailed Case Studies*. Washington, DC: World Bank, Operations Evaluation Department.
- World Bank (2007) *Aid Architecture: an Overview of the Main Trends in Official Development Assistance Flows*. Washington, DC: International Development Association.
- World Commission on Dams (2000) *Dams and Development: A New Framework for Decision-Making*. London: Earthscan, URL (consulted 7 November 2007): <http://www.dams.org/docs/report/wcdreport.pdf>

SEAN CONLIN is currently a visiting senior research fellow at the Department of Anthropology in the University of Sussex. Sean also set up the ConMara Partnership, a development consultancy. During a long career as a social development adviser in the UK's Department for International Development, Sean gained a broad experience of development policy and practice in Africa, Asia and Latin America. Sean also worked in the World Bank, the European Commission and the Organisation for Economic Co-operation and Development. Sean has

Evaluation 14(2)

retained an interest in how social science can inform development policy and practice and the use of social science approaches in the evaluation of results.

RODERICK STIRRAT is Professor of Anthropology and Director of the Centre for Culture, Development and the Environment at the University of Sussex. Besides his academic work, he has also a broad experience of social development in general and evaluation in particular, and he has worked for, among others, the Department for International Development, the African Development Bank and various NGOs.