Own Goals: 
Democracy, Evaluation and Rights in Millennium projects

Saville Kushner 
University of the West of England, UK

INTRODUCTION AND SUMMARY

I will argue that international organizations are caught in the middle of a paradox as they seek to implement international goals in local contexts. It is this: that events and phenomena like schools and educational ideas have what appear to be mutually exclusive properties - that they are both unique to context and comparable with other contexts; they have both a global and a local character. Schools, for example, are both comparable across countries and cultures and unique to their host communities. This is a paradox that afflicts UNICEF, as with other international agencies, as it struggles to resolve the tension between its global accountabilities and its roots in a local (in-country, often municipal) action-base with obligations to children, families and communities.

One of the implications of this paradox is that we have to think flexibly about what counts as a standard or as a goal, allowing them to be determined locally – to reflect the priorities and preferences of citizens and communities – and globally – to reflect international agreements and advances in internationally-recognised moralities. In respect of evaluation I will show how thinking simultaneously of democracy and rights in evaluation – in fact, as determinants of evaluation design – allows for the mediation of the global and the local, especially in respect of MDGs. Though rights and democracy are frequently thought of by rights theorists as being in tension, program evaluation allows for their unification. I will talk briefly about a right-based approach to evaluation.
DEMOCRATIC EVALUATION AND GOOD GOVERNANCE

Evaluators have rehearsed proposals for democratic evaluation for many years. For some, the search for singular calculations of the productivity of a program was secondary to the need to ensure that programs were properly held to public account but also that judgements to be made about programs were not the sole prerogative of a social or political or administrative elite. Indeed, democratic evaluators have argued that measures of program productivity may well, in themselves, be insufficient to meet the democratic obligations of the evaluator – taking into account the diversity of values in a program, the instability of output measures, program politics and typical conflicts over program purposes.

What is argued for in Democratic Evaluation is such as: open exchange of information across stakeholder groups; evaluation as a space in which power inequalities can be (procedurally) neutralized – i.e. equal treatment for all (mother, manager, minister); the independence and impartiality of the evaluator whose obligation is to address everyone’s dilemmas; free and open publication of evaluation reports; and a recognition of collective responsibility for enhancing public information (MacDonald, 1976, Ryan & DeStefano, 2000 – but see below for a more detailed account). Clearly, key to any democratic evaluation process is participation as a form of direct representation (Ryan & Johnson, 2000).

Notwithstanding the weight of issues considered by its theorists, Democratic Evaluation as a methodology has been more popular than practiced. This may be due to the demands it makes for sophistication in action, together with the fact that it is often unattractive to sponsors who are wary of the loss of contractual control over the evaluation and its publication and jealous of the resource they see themselves as having paid for. It may also be the case that where public evaluation is mostly practiced is in those advanced industrial countries where it is less compelling to make the link between democratic evaluation and good governance – a link that lies at the heart of the methodology.
Perhaps most significant of these factors has been the widespread failure of the link between Democratic Evaluation and governance. As originally conceived, the methodology both made up for a democratic deficit in social programs (e.g. enhancing the accountability of program managers and political sponsors to other program stakeholders) and offered to make the conditions of our democracies transparent by seeing a social program as a microcosm of society at large (i.e. each evaluation is a case study of political society). Democratic structures and assumptions are sufficiently strong in Europe and the US to dilute the political imperative behind Democratic Evaluation.

The imperative behind this link is stronger in international development settings where the absence of ‘good (i.e. democratic) governance’ and the frequent absence of program accountability to stakeholders and citizens is prevalent. Here, Democratic Evaluation would seem to hold promise for those who advocate more open, accountable and responsive forms of governance – at program, municipal and national levels. However, preoccupations over democracy and information rights have permeated only little into evaluation practices in international development (see, for example, Cameron & Ojha, 2007; Segone, 2006). Program evaluation for development is mostly confined to impact assessment and, increasingly, audit and accountability. The search for the elusive unified calculation of quality or program productivity dominates international organizations who are under accountability pressures from their donors to see programs of intervention less as opportunities to engage citizens in deliberations over priorities and more as delivery systems for pre-specified results. The link between democratic evaluation and democratic governance has not been sufficiently well forged.

But democratic evaluation has always been promoted within the context of national political cultures and does not easily transfer to global contexts. It is probably fair to say that democratic evaluation was implicitly modeled on civic democracy in a liberal, Western state, with expectations of social exchange, deliberation and the building of social
consensus. Indeed, Deliberative Democratic Evaluation (House & Howe, 1999) – its most recent expression – emphasizes the kind of social exchange that is almost inconceivable beyond levels of local intimacy in political cultures. Democratic Evaluation functions through social conversation – so it needs to talk in a national ‘language’. The warrant of the democratic evaluator emanates from the democratic aspirations of the society which commissions it – it is an echo of political ambition.

At a level of global action we cannot assume or achieve political singularity – there is no such clear echo. There is no civic context to a global world. Here, we struggle to find the source for that same democratic warrant. The signatories to Millennium Development Goals (MDGs), for example, are too diverse in their political hue, at too diverse stages on democratic journeys to provide a stable base for evaluation mandates placing these goals under critical scrutiny. I will take a detailed look at MDGs a little later, for they stand as a global program and are the test case for looking at the constraints and possibilities of democratic evaluation in global contexts. The challenge for the democratic evaluator is to find a way of grounding a constructive critique of MDGs in what are often referred to as ‘beneficiary communities’ – as a way of transferring ownership to them of the goals.

In this chapter I explore a rights-based approach to evaluation, one that draws for its warrant from international statements of authority – treaties and conventions. A rights-based approach to evaluation is not a substitute for democratic evaluation, it is merely an adaptation that may be appropriate to global programs and to international contexts – a means for bridging awesome distances between international elites and disadvantaged communities. It has the added advantage of a vocabulary that is familiar to development workers and organizations. All contemporary international development programs are required to be ‘rights-based’ (Jonsson, 2003, Black, 2008).

I start the chapter, however, with a paradox that always faces the educational evaluator, but which is heightened in the case of global
action. This is the paradox of what we might think of as ‘situated generalisation’ – that any educational event has Heisenberg-like properties that make it simultaneously idiosyncratic and law-like. How we resolve the paradox in action – in, for example, designing an evaluation – has often grave implications for how we distribute rights.

A GLOBAL PARADOX

Wherever we travel in the world we find classrooms recognisable - instructional arrangements, many small people facing one large person, authority structures and power asymmetries, repressed desires, scrutiny, revelations, learning, judgement, confusions between play and work...we see and readily recognise them all. Under a Baobab tree in Central Africa, in an adobe shack in the Bolivian highlands, in the suburbs of an English city, in the shadow of a Buddhist temple in Thailand and in a tent in a Middle-Eastern desert – teachers teach and pupils learn in ways that make it easy for us to hold educational conversation across cultural boundaries. Classrooms are familiar settings. Partly as a result, the language of pedagogy has become a global language. From all over the world educators come to Europe to reflect on their practice, and our cultural and political differences apparently present no barrier to conversation. I have visited Malaysia only fleetingly, and yet through a series of doctoral supervisions I seem to have come to know Malaysian schooling to the point where I am taken seriously in conversation by educators there. How can it be that we pretend to such cross-cultural knowledge while still cherishing the view that our cultural traditions confirm our uniqueness, make us exotic to each other?

Look again at classrooms. We enter a primary school in our own ‘back yard’ as a parent or as a researcher, and we are acutely aware – especially as parents - that two classrooms next to each other on the same corridor are two different worlds. One, perhaps, has a middle-aged,
energetic teacher with a progressive approach to open education and holding an overriding educational principle that you have to love the children – and from that, all else flows; in the other class, a young, inexperienced teacher made nervous by overwhelming complexity and a fear of children is conservative and disciplinarian, and focuses, not on the children, but on the formal curriculum. The two teachers can as easily talk to each other as walk together to the canteen – but they may well find the distance between their respective value positions too great to overcome. Their classrooms are their invented countries, the gap between them is as wide and as deep as the gap between cultures, and it frustrates attempts at conversation. We know this well from the history of school-based innovation which too often fails through the difficulties innovative teachers have of negotiating their way with colleagues. On this side of the paradox classrooms are unique, far from readily familiar, and the unpredictability of their cultural formation denies the possibility of universal description. Not only this, as pupil cohorts and teachers themselves change from year to year, so the cultural formation of each classroom shifts and changes. The global classroom dissipates into unstable idiosyncrasy.

The same, of course, goes for schools – again, especially primary schools – where the head teacher, as head of a professional ‘family’ sets the cultural tone in ways that are as distinctive as his or her own personality and biography. Schools, too, are both recognisable and different one from another. The social order of schooling is, in one sense, enduring – ‘frozen’ as House (1974) put it – both stable and predictable across time. Schools are places for the reproduction of cultures, metaphorically speaking they are genetically programmed for the continuity of a political strain – and that ‘strain’ is increasingly international. A national curriculum in Malaysia and England teaches the same Mathematics the same Grammar and the same science – the same music, since English schools started teaching World Music. History may be different, but the historiography stays the same. There is no Asian state schooling approach to Science or to calculation. Whatever the hopes of development agencies for ‘child-
friendly schools\(^1\), the global order of schooling is as ‘unfriendly’ to children but friendly to the economic and political elites as is portrayed in the impressive superstructure of the PISA and IAEA\(^2\) studies: relentlessly focused on numeracy and literacy and on the production of an economic workforce. The PISA/IAEA studies compare reading ability in the UK and Thailand!

In the other sense, the order of schooling is local and variable – of course knowledge varies with culture. Curriculum fits into a child’s life and is accommodated and adapted there. All knowledge is personal, schools are instruments of local meaning – it is not useful to talk of ‘all schools’ since their effects, their products, define them and these are both unpredictable and variable. An adult literacy class in that adobe shack in the Bolivian highlands has indigenous men and women playing out the shifting village politics of gender and entertaining their personal dreams of empowerment.

And this is the point of the paradox – that schools and classrooms are identical and different at one and the same time – the PISA studies are both valid and invalid according to the criterion of judgement you choose. Neither is deniable. Here is the central methodological challenge of educational evaluation – how to estimate the worth of an educational activity where we face these contradictions. Do we measure one classroom or school against others – do we, that is to say, define universal standards against which to measure and compare performance based on the assumption that schools and classrooms are comparable – as in the global PISA and IAEA comparative studies? Or do we have to reinvent measurement which is tailor-made for each classroom – do we, that is to say, define standards appropriate to each classroom and each school? Here, indeed, is a challenge for international development

---

\(^1\) http://www.unicef.org/lifeskills/index_7260.html#A%20Framework%20for%20Rights-Based,%20Child-Friendly

\(^2\) Program for International Student Assessment, OECD (www.pisa.oecd.org) describes itself as an “internationally standardized assessment” of pupil performance, currently across 57 countries. *International Association for International Assessment* (IAEA – www.iaea.info)
agencies whose liberalism insists on equal treatment for all – the global expression of rights.

The constant reinvention of measurement was, indeed, the view of Stenhouse (1967) who regarded a ‘standard’ as a set of local agreements over what was worthwhile. By standard he referred to "criteria which lie behind consistent patterns of judgement of the quality and value of the work" (p.70) - i.e. a standard was a procedural principle or a process, not a universal and fixed measure of performance. Importantly, the 'consistency' of these criteria rested upon interaction and communication – they were, in Stenhouse’s terms, based on local judgements and values. Let us look at their implications for the measurement of quality. Stenhouse says:

"When we say that we regard the work of one class as better than that of another, we are not simply judging it to reach a ‘higher’ standard. Such a conception implies a common measure against which both classes can appropriately be assessed, but in fact standards can be qualitatively different. When they are, a comparative assessment is not a matter of measurement but a matter of value-judgement. For example, we may opt for creativity or correctness....Such choices are founded upon conviction rather than demonstration. The sources of standards in school work lie in the teacher's values." (1967:75)

As against this we cannot deny the value global measures of classrooms and schooling have created. Gage (1996), for example, reviews the successes of meta-analysis in the behavioural sciences and the very concept of generalisation across educational contexts. He analyses empirical experience with educational generalisation following Cronbach’s (1975) critique that ‘generalisations decay’ (i.e. over time and location)³. Gage’s issue is this:

³ “At one time a conclusion describes the existing situation well, at a later time it accounts for rather little variance...the half-life of an empirical proposition may be great or small.” P.121
"How often do ecologies studied in the behavioural sciences yield main effects...sufficiently substantial, consistent and enduring to support lasting generalisation?..The important question concerns, not the magnitude of main effects and interaction effects, but their consistency.’ (pp. 10-11)

Consistency rests upon the assumption of universal or global characteristics of classrooms and classroom interactions that allow for ecological, cross-context measurement. Gage points to a range of key educational generalisations which appear to be founded upon such consistency – they take the form of relationships: socio-economic status related to academic achievement; years of schooling to amount of knowledge gained; co-operative learning related to race relations; personalised instruction to achievement levels; behavioural instruction to learning gains. Indeed, so successful have some of these been that they form part of a widely accepted ‘folklore’ of education. Perhaps they do form a set of global (PISA-type) criteria against which we can successfully make measured comparisons of the performance of education systems globally? Perhaps there is hope for a global liberalism in education?

The implications of the paradox are far-reaching, and pose questions at the heart of contemporary ‘school effectiveness’ movements, the ‘new public management’ (NPM) and their global equivalents in education. This new context is based on low-trust accountability, performance management and results-based programming (Norris & Kushner, 2007) and it embraces the global liberalism of international agencies which are all committed to it. NPM is only possible where we are able to control for context since cross-site comparisons and bench-marking are key to its legitimacy and functioning. In this sense, inter-contextual comparison is designed, not for arriving at broad generalizations of the kind suggested above, but for ensuring compliance with policy – i.e. control-for-consistency. For NPM to operate effectively we have to assume the comparability of classrooms and schools – that the few can be measured against the many. Let us take a critical look at the global approach to NPM that is represented by Millennium Development Goals (MDGs) and at their implication for global evaluation.
An international agreement was adopted by 189 nations and signed by 147 heads of state and governments during the UN Millennium Summit in September 2000. The agreement established 8 universal goals which would drive all international development work and which would be realized by the year 2015 (though their baseline was taken as 1990). Each Millennium Development Goal has sub-targets to specify the action-frame, and each target has a series of indicators against which progress can be measured towards is attainment – there are 18 quantifiable targets that are to be measured by 48 indicators. These are the goals:

Goal 1: Eradicate extreme poverty and hunger
Goal 2: Achieve universal primary education
Goal 3: Promote gender equality and empower women
Goal 4: Reduce child mortality
Goal 5: Improve maternal health
Goal 6: Combat HIV/AIDS, malaria and other diseases
Goal 7: Ensure environmental sustainability
Goal 8: Develop a Global Partnership for Development

To give a single example, Goal 1 has one of two targets which is to reduce by half the proportion of people who suffer from hunger, and this target has one of two indicators which is stated as: *Prevalence of Underweight Children Under Five Years of Age*. So: Goal – Target – Indicators. The Goal provides the moral/political aim; the Target provides the strategic focus and the Indicators express the target in quantifiable terms ready for measurement of progress toward the goal. The hierarchy is like this:

---

4 See the official UN web-site: http://www.un.org/millenniumgoals/
5 See web-site http://www.un.org/millenniumgoals/
Agencies within the UN carry mandates for monitoring progress towards these goals, and there are many monitoring systems in place generating volumes of data. The scale of this program is gargantuan – all UN agencies design their programs around these goals; most NGOs are doing so (not all, see Black, 2008) and increasing numbers of governments are signing up to drive their social legislation in relation to them. For an appropriate sense of proportion we have to measure the MDG program against global movements such as the spread of Islam and Catholicism, the Freudian diaspora, the internationalization of Flight Aviation Authority standards, and the spread of Napoleonic Law. Resourcing is vast – that is to say, the resources dedicated to organization and to interventions related specifically to the MDG program and its derivatives. Each MDG has an associated UN agency which carries a mandate for measuring progress in each country towards each of the goals.

**MDGs and accountability:** As an accountability system this is, again, a program of massive proportion. Almost anywhere in the world (certainly in the developing world) the efforts of large numbers of people are being measured in the same way, against the same indicators – often against the same database; governments are setting up MDG monitoring units; national and regional statistical agencies and UN agencies are working on the ever-present dilemmas of data quality, harmonization of MDG indicators and data availability; all international agencies are subject to intensifying results monitoring, and one measure many adopt is to import MDG performance targets as an off-the-shelf solution. The quality of many international development interventions is measured in terms of their claimed success at stimulating progress towards meeting MDGs – no
matter how appropriate or otherwise that may be to the nature and mission of a particular organization – or, indeed, to the field being developed. In terms of our global paradox, MDGs promote the homogenous, universal comparative view. Let us, then, enter into this side of the paradox to review its strengths and weaknesses. Later, we will come back to the other side of the paradox to remind ourselves of the inevitability of localism.

GLOBAL EDUCATIONAL EVALUATION

The accountability aspect of MDGs is supported by a sub-program within the MDGs – Results-Based Management (RBM). This evaluation approach, familiar to many in the public sectors of advanced industrial countries, requires the pre-specification of goals as program objectives, and the subsequent comparative analysis with outcomes. The principles of evaluation through comparison of program outputs with its goals are the centerpiece of the New Public Management and its reliance on performance management. They are designed to import efficiency into programming by enhancing predictability and control, and to define accountability in terms of efficiency and compliance. High value is given to logic modeling as an approach to the design of programs in line with management by results, for logic modeling encourages the belief in a unifying program rationale and a causal mechanism for improving the predictability of outputs. It is assumed to build-in the evaluability of the program. In a global context, it is taken to ensure that programs in different parts of the world can be evaluated in identical and comparable ways so as to compare results. Hence, for example, in UNICEF there is a great deal of interest in comparing the performance of child-friendly schooling across the world’s regions. Much evaluation in development settings is designed to emphasise the universalist side of our paradox.

---

The MDG + RBM approach has its advocates and its critics. One prominent argument in favour has it that pre-specifying results allows for joined-up action leading to synergies across otherwise fragmented agency resources – something we know to be true. For example, the centerpiece of contemporary UN Reform is the integration of its core agencies – such that, for example, a national government might in future deal with one UN office and source technical expertise there (the reform is labeled 'One UN'). MDGs provide the common denominator which makes at least feasible the integration of what, in reality, are diverse organizational cultures. Another rationale is contextual – that international organizations are characterised by geographical distance (e.g. headquarters from field office) and that a common denominator (a Goal, a Result) allows central authorities to predict and control action at the periphery. This further strengthens the universalist side of the paradox.

A prominent critique of the programming-by-results approach suggests that it embodies a confusion between consistency and coherence – i.e. that the price of creating and sustaining a good pattern at the level of global/corporate aims is disorder at the level of action – something we know less about in the development field given the shortage of critical evaluation (see, for example, Baser & Morgan, 2008, who argue that results-based management is, at best, appropriate for stable and predictable contexts: "once the playing field became more uncertain, more informal, more contested, more intangible and more long term, RBM quickly lost traction and relevance" – i.e. RBM tends to be unsuited to development contexts.) The localist side of our paradox is, that is to say, correspondingly underdeveloped.

To take an example here, we might think of the instability of an MDG indicator. MDG2 seeks universal primary education; the target is stated as by ‘2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling’, and the official UN web-site states the indicator as the net enrolment ratio in primary education. The first observation to make is that the indicator does not match the Goal – i.e. the Goal talks about completion whereas the indicator focuses on
enrolment. A key issue in local development settings is that it is easier to put a child on a school register than to ensure that the child attends school on a regular basis. Here, too, we hear an echo of our paradox. Enrolment is a more stable measure and allows for cross-country comparison; completion requires local verification and analysis and is so contextualized as to defy comparison (it was easier to enroll girls under the Taliban than to guarantee attendance). The goal, target and indicator have other unstable aspects when played out at local level where context and idiosyncrasy overwhelm universalist ambition - such as:

- many countries require children to repeat years of schooling, so completion of a course of primary schooling may mean a significant and prejudicial diversity in age and experience;
- what we define as a ‘course of primary schooling’ is, obviously, ambiguous and likely to frustrate comparison, as is the term ‘full’, which varies from country to country and from family to family;
- ‘boys and girls alike’ may stretch comparability to breaking point since the challenge of enrolling and then guaranteeing the attendance of girls is, in many places, significantly higher than that facing boys – and vice-versa in other places;
- Enrolment, attendance and completion can be measured in diverse ways;
- Enrolment, attendance and completion are minimum expectations and contain no indicators as to quality of local educational experience;
- Baselines are key to measuring progress, given that MDGs are measured, almost exclusively, in an interrupted time-series approach. Since countries develop (and regress) in their own pattern, baselines are likely to be incomparable in themselves.

This is not to undermine the millennium aspiration, but to bring into question its articulation as a goal, its evaluation and how we account for context – the other side of the paradox. What I also want to highlight, for this is the theme of the later stages of this chapter, is the impact of universalist comparators on democracy and participation – in short, who
owns these goals? It is often the case that an indicator may be written in ways that are more or less useful to mothers, families and communities, say, and more or less inviting of them to join the development push. For example, a community or municipality seeking to take control of its own educational development will find local comparisons (i.e. with proximal contexts and/or within the same culture and conditions) more meaningful than cross-country comparisons – whereas a national government setting broad policy parameters might find other country data more informative. A community will find ‘attendance’ measures more useful in solving exclusion, just as national governments may find ‘enrolment’ more useful in resolving international accountability demands.

Let me take one other example from the area of early childhood development. How do we measure chronic malnutrition in pursuit of MDG4? In practice, this is measured either by ‘wasting’ (calculating weight for height) or ‘stunting’ (calculating height for age) and the decision which to opt for is made by an agency or a government or, indeed, an agency’s regional office. It is then applied in a uniform way to communities and families and the measures used as a basis for policy advocacy and institutional development. For mothers on the ground, the choice of method is significant, since one (stunting) is more observable to her than the other and allows her to participate in the monitoring process, whereas the other (wasting) requires weighing technology and measurements over time. Choice of one indicator over the other may be consistent with methodological policy and useful in policy contexts, but less useful and less inclusive for a mother. All such indicators are unstable in one way or another.

*Strengths and weaknesses of the universalist project:* Nonetheless, this MDG framework is the glue that holds together many, if not all, international development efforts in such a way that (a) all efforts are bent to internationally agreed priorities, and (b) that all efforts across international agencies and NGOs are comparable and complementary. The strengths of this program are:
1. that its goals have a democratic warrant, in that they were verified by (broadly) elected governments;
2. they represent an international consensus on a moral framework and priorities for development – a unique political accomplishment;
3. they have measurable indicators of progress which gives rise to a transparent accountability system, consolidating the political gain;
4. they do permit country-by-country and region-by-region comparisons (for the purposes, principally, of resource distribution).
5. A final and formidable strength of MDGs is that they reflect in their substantive content the broad range of specialist and sectoral interests in international agencies which allows for immediate engagement with them (e.g. poverty reduction specialists, nutrition experts, education advisers, etc.).

These are undeniable strengths, valued by international aid agencies and made good sense of in action. MDGs also have their weaknesses, of course. Some are familiar from critiques of results-based programmes (see, for example, Franklin, 2008) and of the general case of the New Public Management (Kushner & Norris, opp.cit.) – for example, that pre-specification of results has a tendency to (a) risk-aversion (i.e. setting low-level goals to ensure attainment), (b) resistance to learning and adaptation to context, (c) limited opportunities for participation (priorities are inflexible), (d) centralization of control. Hence, we may consider weaknesses of the MDG program which correspond to the obvious strengths given above and which highlight, once again, the other side of that paradox:

- **MDGs have a democratic warrant:** While qualifying as a democratic warrant, this is possibly the weakest available – i.e. it binds citizens and professional institutions into the preferences of their political elites. Contemporary democratic theory is tending towards ‘deliberative democracy’ – i.e. subsidiarisation of policy with high levels of participation (Dryzek, 2000).
They represent an international consensus on a moral framework: The moral framework derives from the deliberations of international elites, not from those who are intended beneficiaries. In any event, there is a logical and a rights-based inconsistency to the concept of ‘international’ morality – moral systems are always culturally, even locally, embedded. Concentrating the moral centre of gravity at global and international levels has a tendency to leave national and local levels to concentrate on its applications – i.e. on technical aspects of development.

They provide measurable indicators of progress – a transparent accountability system: Quite apart from the instability of both indicators and their measurements (just hinted at in the examples of primary education and chronic malnutrition, above) this ignores the impact of methodological choices on possibilities of participation and inclusion. The more we focus on technical measurement, the less we make programs and interventions accessible (or meaningful) to civil society – the more we ignore the politics of knowledge internal to a country.

They are universal: A universal social context can be little more than a theoretical construct looking for grounding – an answer looking for a question. Many (not all) aspects of poverty, for example, are contextualized in local circumstances and expectations. But MDGs suffer from a key weakness in all goal-based programs – that the price of universal application is generalization to a point of abstraction. If we want a goal to be responsive to context or a concrete guide to action it must be written in such a way as not to appeal to all.

They reflect the range of specialist and sectoral interests in international agencies: Insofar as MDGs are a 15-year program, the structure of the demands they make on attainment has the tendency to reinforce and freeze that same structure on organization, making UN agencies less able to respond to changing
roles and changing contexts (including UN Reform itself). For example, we may have less need for nutrition, pediatric and education specialists, more need for multi-disciplinary specialists in policy-shaping, institutional development and professional change. MDGs, this is to say, potentially distort the very nature of organization in development agencies towards greater responsiveness to global goals than to local contexts.

These weaknesses do not mitigate the strengths and nor do they imply that MDGs are without merit. However, we do have to treat as problematic how MDGs are realized in contexts of local action and cultural diversity – how, while honouring the universalist side of our paradox, they do not violate localism. We do also have to balance the extensive effort and resources put into measuring progress in indicators-based ways towards meeting these goals. A great deal of local experience, diversity and complexity is being lost. The weight of resources is dedicated to measuring progress towards meeting the goals, and so the rhetoric is one of striving and achievement, not reflection and learning. By the year 2015 attention will shift to a justificatory rhetoric for having met or otherwise a particular goal. MDGs have no change strategy specifically attached to them, and insufficient work is being done to develop evidence-based understandings of change. The risk is that we reach 2015 probably with partial success but with little or no understanding of why certain goals were met and others not, of what were the local mechanisms of change.

Above all, with the dominance of universalism, too little work has been done to ensure ownership of the MDG effort in local, civil society (it is too late to ensure socially embedded democratic ownership of the goals – the program is well under way). It would be a betrayal on a universal scale if, having pursued these priorities over a 15 year period and bent all efforts and resources to the task, to discover that the society created by this program is not one found to be desirable by those citizens who are required to live in it. Access to primary education, birthing resources, clean drinking water, the internet, etc. are notable and desirable technical accomplishments of the MDGs – but we are not collecting data on their
social and political spin-offs and consequences – on the kind of society their accomplishment implies.

The final critique of the MDGs, then, is to point out that they make no mention of democracy and governance and so lose an essential grounding in localism. The goals themselves may be pursued through authoritarian, bureaucratic, democratic or any other means. In fact, they represent a set of ends divorced from means, whereas in democratic terms ends can only be justified by means – that, for example, there are threats to democracy in imposing liberal goals through autocratic, paternalistic or coercive procedures. This is, perhaps, a casualty of universalism.

I turn now to the localist view of the paradox with which this chapter started. I want to resolve the question of how we democratise MDGs – for this is still possible to an extent. I will do this by restating the argument for Democratic Evaluation – or, indeed, with its sibling, Rights-Based Evaluation.

DEMOCRATIC EVALUATION

‘What goes round comes round’, as they say. As international agencies intensify their advocacy of citizen rights it is not unlikely that citizens of those countries they work in will come back asking to exercise their rights. One of the places in which we may expect rights to be asserted is over the creation of social programs which are designed to shape new futures for citizens. Women and children are asked to participate in development programs which are frequently designed and managed by others, and to inhabit new futures promoted by these programs. At what point may they assert their right to make judgements about the kinds of futures being envisaged - how such programs are judged, against what criteria and for what purposes? At what point, that is to say, may we expect people and communities in developing countries to demand the right to information about MDG programs?
As we pursue the Millenium Development Goals we envision societies with reasonable life expectancies among women and children; the alleviation of, at least extreme, poverty; broad educational coverage including gender balances; easy access to clean water; etc. From these flow attempts to reduce disease, to enhance medical care, to give children the right to a birth registration, to enrol children in school. These are the technical accomplishments we aim for. What underlies them, however, is a vision of society radically different from that which many developing countries currently enjoy – with different power relations, different institutional structures, cultural adaptations – societies in which democracy and authority bite in different ways.

To take one minor but significant example. A key strategy in the promotion of children’s rights is birth registration – the official acknowledgement of a child’s existence with which that child can then assert their right to health care, education and participation. Of course, what we are also doing is expanding and reshaping the electoral base of a country. In the case, for example, of Andean countries with large indigenous populations, we are contributing to the reshaping of geopolitics by bringing into the citizenry hitherto excluded groups who will enjoy rights previously denied to them – a project with historic implications, as we have seen, in Bolivia which recently elected an Indigenous man (‘originario’) as President. The question of democratic rights concerns who has direct access to judgement about our attempts to induce and manage social change. This, indeed, is the real promise of development interventions. The democratic evaluator asks, ‘who has the right to know about these things?’ and concludes that all do.

In the advanced industrial countries of Europe, Australasia, the USA there is a lively debate around the merits of Democratic Evaluation – an approach that recognises that program evaluation is a form of (at best, impartial) political action. This is not to say that evaluators are party political activists, but that we recognise that knowledge and information are exchanged and applied in political contexts, and that this demands
that evaluation has a political strategy. Just as we live in an age that is broadly sceptical about the possibility of value-free social science, so we are developing a scepticism about politics-free social science. Insofar as we recognise that while we are conducting evaluation we are acting politically, we look around for appropriate guides to action and to an appropriate warrant. Why? Because it does not take long into a program evaluation that evaluators often are challenged to justify their actions – for example, in providing or withholding knowledge to one or more groups.

One of the early solutions to the dilemma was to link evaluation to democratic precepts. The proposal was not to assert representative democracy in evaluation – for example, that evaluators become or serve exclusively elected representatives – but that what stands as due democratic process should govern the conduct and the transactions of evaluation. In fact, in the earliest expression of Democratic Evaluation MacDonald (1976) specifically excluded the definition of evaluation as a privileged service to elected representatives (he labelled this 'Bureaucratic Evaluation').

The kind of democracy envisaged by evaluation theorists and tested by them in practice is the kind of civic democracy found at municipal levels of society – i.e. the pursuit of collective rights at community level. Here is an approach designed for localism and to combat universalism. Civic democracy is intimate, often face-to-face, and so involves argument and direct exchange. Since civic relations are so close, it is essential to make decisions by negotiation, and this goes for evaluators, too. What the evaluator is prepared or free to negotiate will vary according to their confidence and to their contract. Another implication of the civic model to democratic evaluation is the role of common sense. Science itself is expensive, and it is rarely available to municipal-level actors – but, in any event, local politicians and administrators live close to their constituents and have to rely on personal persuasion more than on scientific demonstration. Argument and persuasion, therefore, tend to happen on the basis of day-to-day language and concepts, and so it is in democratic
evaluation. The democratic evaluator tries to collect data and report it in ways which reflect how programme people think and talk – ideally, in the language of persuasion rather than proof.

But the main implication of this civic politics understanding of Democratic Evaluation is that, as in municipal governance, political structures are defined less as sites for the determination and dissemination of policy and more as sites which allow for argumentation and the shaping of policy – there tends to be a shift of emphasis from policy outcomes to policy processes. Those who govern are within touching distance of the governed and so have to negotiate their way. So it is with democratic approaches to evaluation which seek out that same proximity and intimacy. Hence, these are good instruments for resolving paradoxes and dilemmas created for local communities by distant national and international elites. Evaluation is not a process for arriving at unblemished truths (outcomes), but for the interplay of competing values in defining what is a reasonable outcome.

What follow are some basic dimensions of democratic evaluation practice around which there has grown some agreement. Look carefully for echoes of localism – for it is only through (institutionally) intimate forms of interaction that these can be feasibly guides to action:

- All people own the data over their own lives and work – evaluators need permission to use it. Some people use protocols – ethical agreements – which govern the conduct of the evaluation and which are explicit about the rights of all respondents to negotiate their data;
- Key evaluation questions and methodological principles - like ethics procedures – should be negotiated and agreed with programme stakeholders.

---

7 I am aware that I use the term ‘stakeholder’ frequently and I now need to say that the term can be tendentious. It too easily implies, for example, that all who may be termed ‘stakeholder’ have equal status in evaluation. As House has previously reminded us, not all stakeholders are equal and we may have to do something more than treat them all equally. A government minister is a stakeholder in a social program just as may be a pupil or a nurse – but is more skilled, robust and better positioned to advocate for themselves than they are.
• Programme stakeholders, including programme managers, programme workers and recipients, have the right to contribute to the criteria against which their work is to be judged;
• Evaluation should adopt methodologies which permit and which do not deter participation and exclusion, and which capture the lived experience of program stakeholders;
• Evaluators have no warrant to grant privileged access to any stakeholder – they must serve all equally. For example, that one party has the resources and the warrant to sponsor an evaluation gives them no proprietary rights over it – they take their place along with all others as respondents and stakeholders;
• Programme stakeholders have the right to know about the views and interests of those who influence their work in whatever way. Ministers and mothers each have the right to know of the others insofar as they one way or another impact on their lives and aspirations. Evaluators have the obligation to support open information exchange;
• Evaluation should be inclusive, seeking out a range of views including those which may be highly valued as well as views which may be controversial – disreputable, even. Evaluators have no warrant to censor or select views;
• Evaluators have no warrant to make their own judgement – their job is to help articulate and to feed into other people’s judgements. This means that the evaluator must be, at the least reserved, about making recommendations. (This is most probably subject to less of a consensus than the others.)

Together, these define the localist project for the evaluator – the obligation to inform civil society and its professional institutions. Here are the links between democratic evaluation and good governance.

**RIGHTS-BASED EVALUATION**
We can invoke the concept of ‘information poverty’ – i.e. a parallel to material poverty with the same effect of reducing the capacity for self-determination. Under information poverty people have insufficient information about social arrangements that determine their lives to be able to make decisions and judgements about them. This is the case with MDGs. For the most part, monitoring and evaluation of MDG progress is locked in reports that are fed back to international organizations, government and the administrative system – i.e. the ‘information wealthy’. There are few established mechanisms for communities – or even their representatives – to receive that information and act upon it. Think, for example, of professional practitioners – the nurses, police officers, social workers and teachers who generate the results international agencies claim. Where they do receive feedback on programs that envelope them it will tend to be framed as advocacy and may not often represent their immediate dilemmas. (This has often been represented in the development field as the denial of local knowledge in favour of Western ‘science’ – see Long & Long, 1992 for one of the early expressions of this.)

In any programme of intervention those who are expected to live with the consequences of the program – i.e. living them or working with them - have the right to know about it: its accomplishments, its flaws, its logic and its history. Just as with material poverty, the solution to information poverty is redistribution – i.e. from the ‘information wealthy’ to the ‘information poor’. This has a range of procedural implications: providing access to the framing and sponsoring of enquiries; making reports publicly available; adopting methodologies which represent the real dilemmas people live with; sharing information across stakeholder groups in such a way as to allow for informed public accountability. This is about public conversation, more than management information.

The principle of rights provides the conceptual leverage for redistribution and for arriving at these procedural imperatives: some people have rights over information; others have obligations to provide it. Evaluation itself,
as the generator and broker of information, is the redistributive mechanism.

The international commitment to rights-based approaches goes further than the claim people make to individual rights. Indeed, the principal (among many) critique of the rights approach is that aimed at the various UN Declarations of rights: that they individualise action rather than collectivise it (Uvin, 2004, but see the counter argument, e.g. by Howard, 1995). The practice, however, is more complex. The human rights-based approach to programming (HRBP) focuses development interventions on policy-shaping processes, withdrawing from direct, field-based action so as to gain purchase on broader determining factors of poverty and exclusion. The aim is to take on the challenge of structural change in developing societies, redefining poverty, for example, less as an issue of economic pathology and more as the structural failure of society – i.e. a failure of rights and a failure of those in power to meet both their social and economic obligations. A rights-based approach to child protection, for example, might involve direct action to rescue particular children from sexual and economic exploitation, but, importantly, would see collaborative action with municipal authorities to establish child protection commissions as principal duty-bearers, and then to ring-fence budgetary allocations to guarantee their sustainability. In terms of the rights vocabulary, HRBP goes beyond the conventional differentiation of rights into civil/political rights, on the one hand, and economic, social and cultural rights, on the other. This is, at best, an integrating framework.

This, too, is directly translatable into evaluation. Any program evaluation addresses both the politics and the economics of social innovation since the program commands resources as well as the power structures that give rise to the pattern of resource allocation. Within a rights-based approach it becomes imperative, then, to build democratic evaluation processes into local political structures on the principle of inclusion. This implies, for example, public observatories, a political facilitation role for universities, evaluative forums, mobilizing schools to support community evaluative enquiry (of the kind developed in the OECD/ENSI programme,
for example – see Posch, 2001). The other principal objection to human rights – that it tends to subvert democratic structures (see Gould, 2005) – is also resolved in HRBP where the two can interlock.

Rights are not taken in terms of simple individualism – i.e. the conceding of discrete rights to attributed individuals. As part of the shift of development intervention to the social policy sphere the rights agenda addresses collectivist values. The rights activist cannot fail to come up against democratic structures: parliaments and senates, ministries, regulatory bodies, accountability systems and professional institutions. These are the structures that create the space within which rights are realized, and the question of whether rights are realized in individualistic or collective ways is an important debate. Indeed, a rights-based approach would almost inevitably lead to advocating a subsidiarity of policy shaping coinciding with Deliberative Democratic Evaluation (House & Howe, 1999) since, individual or collective, rights may be protected and resourced at international and national levels, but can only be realized at local levels. A rights-based approach to evaluation seeks evaluation providing information in a place, at a time and in a way that supports this local realisation of rights as well as the shaping of the policy umbrella to support it (what Uvin, 2004, describes as the ‘empiricist strategy’ to rights implementation – i.e. piecemeal, locally defined, non-universalist).

In the context of rights-based evaluation, and in the international language of rights, this defines the evaluator as a duty-bearer. Indeed, we might go so far as to say that the evaluator enjoys no rights at all (other than those held in trust for their respondents), but that the role is defined exclusively and uniquely in terms of obligations – obligations, that is, to allow for the redistribution of information goods. The rights-based evaluator in a democratic set-up is the servant of all rights-bearers and is warranted to grant no privileges to any. The agency that commissions an evaluation shares, by complicity, the same set of obligations and can be held, in the same way, to enjoy no privileged rights. Their role is defined as ‘sponsor’ not ‘purchaser’. This goes for those agencies – including the United Nations - which commission evaluations of MDG progress.
Evaluation sensitive to rights issues is an instrument for working on the issue of how to translate universalist rhetorical statements of rights (such as international conventions) into local action, taking account of local complexities. There has to be a dynamic tension between international agreements like MDGs and local preferences, since human rights conventions cannot ignore contexts, but nor can local authorities declare immunity from them. Similarly, we can ignore neither collective nor individual rights; and we have to be able to cope with the reality of people being rights-holders and duty-holders at the same time. Solutions, as always, are case-made at local level and regulated by decent, democratic and fair accountability systems – evaluation provides the resource and the opportunity to theorise solutions at the local level and without it such resolutions are not available. In the end, international agreements and treaties – including MDGs - should be judged for their relevance to local contexts.

CONCLUSION: EVALUATION FOR SOCIAL CONSENSUS

All societies are characterized by fragmentation over purposes and values. Professional practitioners, program managers and ministers, little understanding each other’s challenges and accountabilities, frequently differ over how best to secure social change. The citizen, distant from exchanges between these groups, is rarely even party to them. If ‘participation’ meant sometimes overturning the aspirations of the political and administrative elites it would not happen – or be advocated – as often as it is. There is no need to be sentimental.

Most serious of all are breaches of communication and understanding between practitioner and government, for it is that relationship that is most potent in generating development gains and ensuring that social investment is both utilized and utilized well. Democracy may fall as a result of political corruption, but it can only stand on the basis of well-founded, publicly accountable and responsive social institutions. The most
democratically aspirational government without a consent-based, efficient police service and school system will come to nothing.

Social fragmentation is corrosive of understanding between these two groups. Program evaluation is potentially nurturing of such understanding. Evaluation which is motivated by an ideology of rights and focused on democratic approaches to their deliberation and realization offers a site in which differences can be made transparent, legitimated, discussed and resolved. It is important to say that the resolution of differences and the forging of consensus over, for example, MDG action need not be a permanent state and need not launder fundamental differences of approach and ideal. We are talking about overlapping consensus (Rawls, 1996; Umphres, 2008) based on common moral principles, and temporary agreements focused on immediate priorities – both of which allow for continuing differences of purpose and value, but allow for immediate and concerted action. The mechanism for arriving at an overlapping/temporary consensus through informed argumentation is evaluation. House & Howe (opp.cit.), for example lift program evaluation out of the realms of technical action and into the realms of the political.

In terms of our opening paradox a rights-based approach to evaluation, set within a commitment to democratic structures and processes, helps to shape that mechanism. The challenge for MDGs is to have them owned at the level of community – i.e. by those who have to live with their consequences. That ownership has to be negotiated and can neither be assumed nor imposed, for the question of whether programs of intervention which are motivated by MDGs are appropriate or desirable is entirely empirical and tested in local action. The only legitimate party to facilitate that negotiation for ownership – legitimate by virtue of being independent and impartial – is the evaluator. Whose rights is s/he championing?
REFERENCES


