The administrative capture of evaluation

Saville Kushner

Abstract: *The past 30 years has seen a gradual internalisation of program and policy evaluation into the machinery of government. This happens either through the evaluation function being taken on by administrators themselves, or by those administrators stipulating and proscribing how the evaluation is conducted and for what purposes. This poses important questions for democracy and for the citizen. This article examines the consequences of this "capture" by the administrative system by looking at two examples: one, taken from the field of international development; the other from a government agency commissioning of an evaluation of one of their programs.*

Keywords: *Evaluation independence, program evaluation, politics*

INTRODUCTION

In his paper on *"the executive investment in evaluation"* MacDonald (1981) raised the question as to what value the administrative system has in commissioning evaluations. These are potentially dangerous activities for a political bureaucracy. Professional observers (evaluators) who often place a high value on their independence are given intimate access to what are often confidential sources of information – and they have the capacity to publish. Of course, evaluators provide what can be useful information, but this does not always outweigh the political risks. But more importantly, as MacDonald says, only the rationalist thinks that a social program is the means to political ends (which makes evaluation useful in a formative and democratic sense). The reality often is that a program is the end itself – i.e. it serves political purposes with no further implications beyond its closing date – in which case the evaluation is just *"*symbolic…an essential feature of the interior décor of any self-respecting agency*"*.

Here, then, are MacDonald"s principal reasons why bureaucracies do sponsor evaluation – the psychological challenge of reading this is to keep a firm balance between cynicism and realism:

- No choice (legislation, regulation, orders, pressure, earmarked funds).
- The only alternative may be an unacceptable level of consultation with, and participation by organised stakeholders with conflicting agendas.
- It keeps potential or proven critics harmlessly occupied.
- The presence of specialist evaluators discourages, or at least postpones, non-specialist and more penetrating scrutiny of the program.
- It buys time: time for the executive to defuse the concern that made the program investment necessary.
- The executive may require, especially in the early phases of the program, a reasonable level of implementation, and the evaluation presence will discourage subversive behaviour within the funded constituency.
- It signals to those who harbour doubts, the executive"s willingness to learn, concern for effectiveness, openness to evidence.
- It is safe, sure to be rendered innocuous by a combination of epistemological frailty, social unpopularity and evaluator self-interest.
- The best reason of all – somebody else may fund an evaluation of their program.

He goes on to show how the executive (administrator) can protect themselves against the risks of commissioning an evaluation – including, for example, *"*delay the evaluation by the escalation of demand upon it…*"*. MacDonald is not, here, arguing that evaluation should not serve the administrative system – but that it should not serve the administrative system *exclusively* – and that while providing the service, it is as well for the evaluator to know the intentions behind their commissioning.

At the time MacDonald was writing evaluation was an expanding, but little elaborated practice – and it was mostly located in universities and with independent academics. MacDonald was talking about *external* evaluation. Since then, there has been a global movement for the internalisation of evaluation, bringing evaluation into the administrative system as an in-house activity. In MacDonald"s analysis this controls for all the risks and allows for an exclusive focus on the benefits – however they are framed. There are two principal means of this internalisation: official agencies employing their own evaluators; official agencies

commissioning external evaluation as a contractual service with stipulated conditions and defined by the management information needs of the agency.

What I want to do in this article is to look at two of the effects of this internalisation: one, in the field of international development, where we will see how administrative concerns with "results-based management" overwhelm participatory and democratic engagement through evaluation; the other, closer to home is a case of the government contracting of evaluation in the UK, where we how internalisation of evaluation has led to tight frameworks of stipulation which pre-define evaluation methodology and limit its purposes – i.e. making evaluation less responsive and democratic. Clearly, what I am concerned with here is the utility of evaluation for democracy. First, however, we need to see evaluation in the expansive, democratic terms in which it was rapidly being conceived in its infancy. In recent years, as evaluation has been internalised as a function of government administration and reform, its possible range of forms and approaches has been increasingly restricted to post-hoc measurements of program productivity. This disguises the fact that the 25 years following the emergence of program evaluation as a discrete practice and discipline saw evaluation portrayed in a wide range of functional and purposeful identities. These are reflected in the wide range of titles given to distinct approaches – each of which (in the following examples) suggest their distinctive character[1]:

Democratic Evaluation (MacDonald)
Goal-free evaluation (Scriven)
Responsive Evaluation (Stake)
Illuminative Evaluation (Parlett and Hamilton)
Evaluator as public scientist (Cronbach)
Naturalistic evaluation (Guba)
Empowerment Evaluation (Fetterman)
Context Input Process Product evaluation (CIPP- Stufflebeam)
Dialogic Evaluation (Greene)
Affirmative evaluation (Mertens)
Utilisation-focused evaluation (Patton)
Personalised evaluation (Kushner)
Deliberative Democratic evaluation (House)

It is worth noting, in passing, that almost all of these authors were writing in the context of *educational* evaluation – and one of my examples below is specific to that field. This is no accident. For the most part, the foundations of program evaluation were laid in the field of education – more precisely, schooling. This was partly since this was the laboratory in which government conducted early experiments with commissioning evaluation – see (Norris, 1990; House, 1993)

To appreciate the full extent of how program evaluation has been conceived outside of the restricted definitions given by administrative systems, Cronbach and his "associates" (1980) famously preface their seminal text on program evaluation with 95 "Theses" about evaluation – statements of value and principle – a manifesto, of sorts. This was written at the same time that MacDonald was speculating about the use

1 See (House, 1980) for a classification system of evaluation approaches and a trend analysis of the emergence of evaluation; and Alkin (2004) for the "evolutionary tree' of evaluation.

of evaluation by the bureaucracy, and a similar concern underpins Cronbach"s offering. He talked of the evaluator as "public scientist". The first of the 95 Theses says, simply:

*"*evaluation is the means by which society learns about itself"

The sense in which this was offered is that each evaluation of a program is a case study of society, in that all programs represent, in miniature, the society which is its host. It has social and power structures, culture, roles and rituals, an economy and a theory of justice. Whether or not this is acknowledged by the evaluator – whether or not her contract allows for this, it is what is happening and it raises many important question that have been widely explored subsequently about the democratic status of programs and their evaluation – for example, who has the right to know about programs and their work? This conception of evaluation is far from

With this and similar questions in mind I will look, now, at my two examples of the impact of the internalisation of evaluation. The first is the state of evaluation in international development, and it starts with a personal story.

1. EXAMPLE ONE: "INSIDE THE TENT" OF DEVELOPMENT EVALUATION

When I was being interviewed for the UNICEF post I eventually held as Regional Officer for Monitoring and Evaluation I found myself in a teleconference covering four countries – Brazil, Panama, Nicaragua and the USA. For some reason it seemed important to keep asking to which country I was talking at any one moment. The most insightful question came to me from New York. *"*Interesting CV – very creative – what a range. But for 25 years you"ve been outside bureaucracies looking in; now you"re proposing to work inside the biggest bureaucracy in the world, looking out. How will you handle the psychological shift?" I merely responded that I had no idea but acknowledged the issue – that this was an empirical matter – I had to resolve the well-identified tension as I confronted it. I was praised for my honesty.

I did not, however, respond to the full implications of the wise question which I subsequently confronted. "What happens," was the real question, "when evaluation comes inside the tent?"

The United Nations is a fragmented organisation. It has many functional agencies which separate out its efforts, for example, on humanitarian relief, cultural development, HIV/AIDS, children"s rights, climate change, poverty reduction, population movement and food production and distribution – whereas the lives of those who live in the Southern hemisphere experience many or most of these things in integrated ways. Climate change, access to schools, migration, food scarcity, disease – all can be experienced as a package. No wonder there is great pressure inside the UN – and among its donors – for harmonisation, standardisation and integration. The flagship reform program is called *One UN* – and you can see why.

A similar fragmentation exists within agencies, too. In my Regional Office we had specialist advisors on paediatrics, vaccination programs, education, public policy, child protection and others – and their operational activities were mostly separate. In fact, there was some resistance to integrating them. *"*I am employed as a

paediatrician!*"* protested one such Advisor when I tried to suggest they were all, in common, change managers. Perfectly reasonable, but tending to the dysfunctional in the face of integrated social problems.

Perhaps the most serious sense of fragmentation comes with geography – now recall that teleconference. UNICEF Headquarters in New York, for example, looks out at vast distances between itself and its seven Regional Offices; from those to its more-than-100 Country Offices; and from those to myriad field offices and operational sites. The question for the organisation is how to interpret that distance? This can be done in one of two ways: as a problem of understanding (learning from diversity), or as a problem of control (reduction of discrepancy). This is a conventional centre-periphery issue. Headquarters has two preoccupations: how to learn from its ground operations, and how to ensure that its policies are followed with a consistency that allows UNICEF to speak with a coherent voice. Both are real and unavoidable, and both are, in varying degrees, taken on board. Indeed, UNICEF is known as the most devolved of the UN core agencies, with its Country Representatives enjoying more autonomy than their counterparts in, for example, the United Nations Development Program (UNDP) – the largest of the UN agencies - which responds to the geographical issue by having a more hierarchical form of organisation. The UN (if it is possible to talk of a single UN) is, in management-speak, a highly decoupled and heterogeneous organisation.

But there is currently a shift of power back to the centre under the auspices of the UN reform process. Indeed, like all UN agencies, UNICEF has to comply with Results-Based Management, a version of the New Public Management (pre-specified results as a basis for low-trust accountability and performance-based resourcing). This, as we know from working in public sectors in the North, is a methodology for centralising power and authority in an organisation. As I left UNICEF, Headquarters was implementing a new policy that changed the role of (powerful) Regional Directors from advice and facilitation, to advice and *oversight* – they were expected to be more demanding of certain kinds of results, responsive to global comparisons of progress towards meeting the Millennium Development Goals. Devolution remains, but is being overlaid with a concern to ensure greater levels of homogeneity. The balance is, I think, shifting as the Centre"s confidence (under pressure from Donor countries) in pushing global agendas (such as gender equality, child violence, child survival measures, avian flue preparations) overwhelms its desire to learn from the diversity of its country operations.

One symptom of such a shift is the call for "harmonisation" and "standardisation". There are three principal reasons for this, two obvious and one more subtle: one obvious one is that there are often alternative ways of measuring social pathologies and it is often important to agree on a single one. Malnutrition is most generally measured by two alternative surrogates: stunting or wasting. Stunting is measuring height against age; wasting measures weight against height. They result in very different demographics. They can, of course, be used together (and should be as mutually moderating measures) but are also sometimes used as options to arrive at the most favourable political statistic under the circumstances. To prevent this, measurement harmonisation is promoted as a way of regulating the politicisation of information. The other, more prominent and publicised example, is that of measures of HIV/AIDS prevalence in a given population. In the UK, for example, we long ago stopped making extrapolations from statistics drawn from testing women in pre-natal clinics – a useful "captive" sample. The biases are obvious (we were, for example, measuring only sexually active women and women who self-present) and statistics early on were scarily inflated. In Africa such methodologies persisted for many years later and inflated estimates of the disease prevalence in those countries – remember figures showing one-third of populations infected? Modifying methodologies was resisted, not least because high

rates of prevalence stimulated greater inflows of "exceptionalist" funding[2].

The other obvious reason for harmonisation is that there is a pressing need for the bureaucrat to compare countries and regions with each other. This is part of the global accountability regime that protects this historically unique phenomenon – the MDGs. Notwithstanding the question of global realities overwhelming contexts and cultures, the fact remains that the UN"s oversight of global accountabilities – in rights, climate change, conflict and violence, for example – has been a force for moral advancement and the reduction of chaos (Black, 2008).

The subtle reason is what concerns me here, and this is to do with the "administrative capture". Administrators do not have the luxury we evaluators have of documenting, theorising and commenting upon complexities – they have to work within them and resolve them – that was the principal cultural shift I experienced. And they have to resolve them in a context of intense accountability to donors for measurable, attributable impact. Circles have to be squared – the desire for aesthetic elegance of input-causal process-output-impact is overwhelming.

Sometimes that elegance has correspondence with reality – especially where problems are accessible to technical solutions: percentage of children vaccinated, number of births registered, coverage of impregnated mosquito nets, intake of iodised salt, proximity to clean water – all have somewhat attributable and roughly measurable impact on health and rights. So much more of international development does not – and the more so since the beginning of another shift: from field-based operations to policy advocacy, where impact is uncertain, inferential at best and contextualised. Here, evaluation as the analysis of context is essential – as is the use of evaluation to reconnect policy advocates with community realities. The further we move away from technical solutions to highly specific problems (like vaccination) the more we enter into uncertainty over what a result is, what it means to different stakeholders, where it comes from, what it represents in terms of quality, and, most importantly, how dependent it is on contexts of action. "In the end, there is no substitute for careful evaluation of the chain of evidence and reasoning by people who have the experience and expertise in the field", argued Deaton (2008) in questioning the over-extension of experimental approaches which may be useful for narrowly-focused interventions, but which fail to represent the complexity of social programs.

Here, we need evaluation for what are too often considered ancillary functions: learning, information exchange, description, deliberation, making context transparent. Where causality is not clear or accessible we need evaluation to reveal contingency – less a matter of how A gives rise to B and more a question of how A and B have a dynamic relationship with each other, but also with X,Y and Z. Picciotto (2007) argues strongly for evaluation to place development interventions in their full economic and political contexts – to lift them out of a myopic focus on the intervention itself and its objectives.

---

2 See for example: (i) "AIDS cases drop, but bad data to blame' Washington Post, 20.11.2007 written by Maria Cheng; (ii) Chin, J. (2007) *The AIDS Pandemic: the Collision of Epidemiology with Political Correctness*, Oxford: Radcliffe; (iii) Glynn J.R, Buve A, Carael M, Musonda R, Kahindo M, Macauley I, Hawken M, Tembo F, Zekeng L. (2000) "A new method of estimating HIV prevalence in men using antenatal clinic sentinel surveillance' paper to the *International Conference on AIDS*. 2000 Jul 9-14; 13: abstract no. ThOrC669; (iv) McGarrigle, C.A, Cliffe, A, Copas, A.J, Mercer, C.H, DeAngelis, D, Fenton, K.A, Evans, B.G, Johnson, A.M, and Gill, O.N. (2006) "Estimating adult HIV prevalence in the UK in 2003: the direct method of estimation' in *British Medical Journal*, 82(3), pp.78-86.

The difficulty we face is that the administrative capture of evaluation has a tendency to depreciate these things in favour of supposedly core functions for which the administrator is accountable: results measurement. As one senior UNICEF official put it to me, "monitoring tells us whether we are moving in the right direction; evaluation tells us when we've arrived". Neat – but fallacious. Since MDGs define universal destination points we can see the rationality behind the desire to harmonise and standardise evaluation. Added to this is a desire to raise and sustain levels of competence and quality in evaluation practice – to protect hard-won gains in capacity development by enshrining them in standards and insisting upon a commitment to learn from "best practices".

The problem is the tension between standardisation and context – the former denies the latter. The urgency to ensure that development policies are implemented and impacts assured has a tendency to displace the democratic functions of evaluation: stimulating public debate and critique; making intervention programs transparent; facilitating information exchange; offering a voice to the citizenry; providing pubic evidence as a basis for conversation between citizenry and official agency. These functions can only be realised in context – these people in this place talking to each other about local concerns at this particular moment. As I have suggested elsewhere, Democratic Evaluation was conceived in the image of municipal, civic politics (Kushner, 2009). At the municipal level people know each other, have face-to-face conversations, there are few resources for sophisticated science, accountability is founded on a sense of relationship, politics are founded on local knowledge and concern. All of these are threatened by harmonisation. Standardisation is the enemy of context.

It is not that this is not acknowledged – certainly in UNICEF. There is a lively debate about democracy and evaluation. The problem is that this has to be seen in parallel with an unavoidable commitment to Results-Based Management, and courageous attempts are made to reconcile the two – to meet both demands together (Franklin, 2008). These are almost certainly irreconcilable, and merely emphasise the difficulty of the task faced by administrators who have to juggle with competing orthodoxies. This is unfair to them and to the citizenry. Democracy is not best seen as a responsibility of administrators.

2. EXAMPLE TWO: "REASONABLE PEOPLE" IN UNREASONABLE CIRCUMSTANCES

The next example comes from the UK and concerns the commissioning of evaluation by a government agency. In the early 1970s the Rothschild Report (1971) was published in the UK setting the terms for the government contracting of research. This defined the relationship between government and researcher as on of "customer and contractor". The customer (government) sets the terms and conditions, the contractor (researcher) complies with those terms under the requirements of a contract. It essentially defined the commissioning of evaluation as an economic transaction – as the "terms of trade" (Pettigrew, 1994). This was the key driver in the gradual internalisation of research and evaluation. It promoted the concept of the "informed customer" – i.e. the administrator who knew enough of research and evaluation to be able to closely specify its conduct.

The subsequent and fairly rapid shift from the commissioning of external evaluation to the internalisation of enquiry was documented widely at the time by evaluators concerned at the control of intellectual and critical enquiry (Pettigrew, 1994, Norris, 1995, Simons, 1995). More recently, concerns over the control of public

knowledge have been embraced by a broader range of concerns over what has come to be called The New Public Management – performance contracting and the centralisation of control (Kushner & Norris, 2007) – including at the level of constitutional change in society (Powers, 1999; O"Neill, 2002). Such discussions have surfaced rarely in development evaluation: this is a democratic issue suppressed by the assumed urgency of action.

What follows is a memo I wrote which eventually led to the British Educational Research Association reviewing its advice to its members on negotiating contracts with official agencies. This documents the detail of the sorts of controls over knowledge that arise when the administrative system assumes responsibility for defining the terms of knowledge generation. It relates to a specification for evaluation issued by a government agency responsible for a central aspect of education.

> *For those seeking to contribute to pedagogical and educational development opportunities to work with the [AGENCY] are of utmost importance. However, the contractual conditions of funding have become so onerous as to be unacceptable. These conditions raise serious questions about the independence and the ethical basis of enquiry sponsored by the Agency, but more worryingly, threaten to compromise the quality of information available to the Agency itself. To deal with each:*
>
> *(1) Funding conditions: the [AGENCY] imposes such reporting strictures on what are frequently short-term projects as to compromise the capacity to engage in serious enquiry and to report proper insights. A recent tender specification for a 10-month contract required fortnightly progress reports, 'ongoing dialogue', and the submission of 6 draft/final reports together with an unspecified number of meetings to discuss the contract and a separate set of three meetings to discuss report drafts as well as an unspecified number of meetings of a 'Steering Group'. It appears that 'contractors' also have to submit and negotiate an 'Exit Plan' within one month of commencement of the enquiry.*
>
> *(2) Ethics and independence: The Agency imposes unusually high levels - even for official agencies - of control over and vetting of reports which are subject to [AGENCY] quality assurance procedures - including for their 'tone'. Extraordinarily it insists upon scrutinising and vetting correspondence to external parties as well as vetting research instruments and reports. Draft reports have to be 'agreed' by the Agency prior to completion as Final Reports. The [AGENCY] reserves for itself the identification of those with the right to receive reports ('audiences'). Not only does the [AGENCY] insist on retaining Intellectual Property Rights - at best a confusion of 'purchaser' with 'sponsor', at worst an abuse of privilege - it also insists on ownership of personal and confidential data which, contractually, must be sent to the agency. This can only have the effect of suppressing critical comment by teachers and teacher-educators who would be advised to withhold views that may be questioning of the [AGENCY], for example, thereby denying the Agency of an important source of views. Quite apart from this, it is not at all clear how an agency can 'own' the personal data of anyone.*
>
> *It is not clear on what basis the [AGENCY] assumes such high levels of control over reports which are paid for by public money and which will frequently hold information of great interest and importance for educational professionals and the parents of schoolchildren. Why do these reports 'belong' to the [AGENCY] more than they 'belong' to teachers and parent-governors, for example? In interaction*

individual employees of the Agency tend to insist that powers of suppression will only be used reasonably and rarely. While individual employees of the [agency] are reasonable and honourable people, 'reasonableness' and 'honour' are not descriptors of a legal entity, nor guarantees of reasonable behaviour. When we vest in legal entities (not its employees) such extensive control over knowledge we create the potential for 'closed society' processes. If the gradual elimination of public rights to knowledge about teaching does not look like a [AGENCY] contract, what else might it look like?

What do we read in the set of arrangements described here? From the Agency"s point of view there is a concern with the quality of the enquiry. It is not uncommon, in recent times in the UK, for such an agency to insist on vetting all research instruments – interview protocols, survey questionnaires, for example, and an explanation given is that the rapid increase in the number of university academics and private consultants competing for evaluation contracts – all seeking income to pay for their salaries – has led to unpredictable levels of methodological quality. Agency"s like the one whose procedures are described here are pushed into a quality regulation role.

Associated with this is the need among Agency staff to protect the reputation of the organization – not least because they have their own accountabilities to meet. The professional commitment of the administrator is not first to the plurality of values, the integrity of methodology or the public right to know – it is to the immediate needs of their political masters. Bureaucracy, by necessity, is a compliant service culture.

3. CONCLUSION

There are two principal lessons to learn from these stories of the internalization of evaluation. One is that the evaluator is denied the opportunity to fashion her evaluation in responsive terms – i.e. in deliberative democratic terms. Early theorists of evaluation in their search for appropriate roles for evaluation defined it less in methodological terms ad more in terms of evaluation as a unique social and political *space* within which certain interactions could happen that could not happen elsewhere or in other social spaces. To enjoy that capacity to define that space (in fact, as ethical space) evaluation has to occupy an independent position. This defines independence not as a privileged role for the evaluator as a political actor, but as a guarantee to the citizen that there is a political (public accountability) process that is afoot and to which they have access. To internalize evaluation is to deny open access to it.

The other lesson is that internalizing evaluation exposes the administrator to public scrutiny and accountability in ways that, in the end, are unreasonable and unsustainable. The administrator is, by definition, not competent to account for the methodological decisions they make and for the methodological and political oversight that follows from their contractual control of evaluation. This is not helpful to the administrator, it presents serious obstacles to the evaluator as we have seen, and it denies the citizen their opportunity to hold decision makers to account. The citizen has no accountability access to the administrator.

If society is to "learn about itself" it is to the relationship between citizen and official administrator that we should look, for the administrative system now enjoys very broad ownership of the means of such learning. It was not designed for this responsibility and, understandably, it does not carry the responsibility

well. Of course, there is no means by which we can, even in a democracy, simply make adjustments to that relationship, for its roots lie in constitutional arrangements – even if contemporary realities have long made that constitution outdated and barely relevant.

However, there are measures we can and should take as evaluators to improve matters. Certainly, professional associations (Psychological, Educational, Sociological and Evaluation Associations) should be entering into dialogue with official sponsors to commence a public dialogue about public information and the obligations of government in this respect. But we should also rely more than we often do on informal agreements over the ethics of evaluation. An Ethics Protocol governing access to data and release (publication) of data, conditions of confidentiality and negotiation of data can serve as a basis for agreement over the fine detail of how to make information public. For example, a clause in such an agreement affirming that individuals own data over their own lives and work carries, at least moral authority when it comes to any attempt to suppress an evaluation report. Such agreements can be negotiated with the sponsor under conditions less pressurized and less proscribed than in the process of commissioning and contract determination. Contracts often displace the values and preferences of sponsors, too, and such agreements can be satisfying to all parties insofar as they manage to express common values.

Ethics agreements have an additional benefit. They modify, to some extent, the economic relationship between citizen, evaluator and administrator/sponsor. Economic relationships (as Pettigrew put it earlier – "terms of trade") are not good vehicles for ethical practices and values-driven interactions. Indeed, we have seen in recent times the catastrophic failure of economic relations where more than a decade of being persuaded to call ourselves "clients", "consumers", "human resources" or "stakeholders" led us into a profound economic crisis – for which contract cultures and poor administrative and regulatory systems were largely to blame. There ought to be considerable moral authority behind the argument that official sponsors of evaluation and regulation play a more humble and democratically responsive role.

BIBLIGRAPHY

Alkin, M., ed. (2004), *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage Publications.

Black, M. (2008), *The No-Nonsense Guide to the United Nations.* Oxford: New Internationalist.

Cronbach, L.J. *et al* (1980), *Towards Reform of Program Evaluation*. San Francisco: Jossey-Bass Publishers.

Deaton, A. (2008), "Randomisation in the tropics, and the search for the elusive key to economic development", *The Keynes Lecture*, London: The British Academy, October 9th.

Franklin, T. (2008), Reaching the Millennium Development Goals: Equality and justice as well as results, *Development in Practice*, *18* (3): 420–423.

House, E. (1979), "Technology versus craft: a ten year perspective on innovation", *Journal of Curriculum Studies*, 11:1: 1-15.

House, E. (1980), *Evaluating with Validity*. Thousand Oaks, CA: Sage.

House, E. (1993), *Professional Evaluation*. Thousand Oaks, CA: Sage.

Kushner, S. & Noris, N., eds. (2007), *Dilemmas of Engagement: Evaluation and the New Public Management*. Oxford: Elsevier.

Kushner, S. (2009), "Own goals: evaluation and accountability in millennium development planning", in K.E. Ryan & j.b. Cousins, eds., *Sage International Handbook on Educational Evaluation*. Thousand Oaks: Sage.

MacDonald, B. (1976), "Evaluation and the control of education", in D. Tawney, ed., *Curriculum Evaluation Today: Trends and Implications*. London: Macmillan.

MacDonald, B. (1981), "Mandarins and Lemons. The executive investment in program evaluation", paper presented at *Annual Meeting of the American Educational Research Association*, Los Angeles, April 13-17.

Marsden, D. (2003), "Rights, culture and contested modernities", keynote presentation to the *INTRAC 5th International Evaluation Conference*, Netherlands, April.

Norris, N. (1990), *Understanding Educational Evaluation*. London: Kogan Page.

Norris, N. (1995), "Contracts, control and evaluation", *Journal of Education Policy*, 10(3): 271-285.

O'Neill, O. (2002), *A Question of Trust*. Cambridge: Cambridge University Press.

Pettigrew, M. (1994), "Coming to terms with research: the contract business", in B. Halpin & B. Troyna, eds., *Researching Education Policy: Ethical and Methodological Issue.* London: Falmer.

Picciotto, R (2007), "The new environment for development evaluation", *American Journal of Evaluation*, 28(4): 509-521.

Powers, M. (1999), *The Audit Society: Rituals of Verification*. Oxford: OUP.

Rothschild Report (1971), Report of *The organization and management of Government R&D*, *Cabinet Office, A Framework for Government Research and Development*, London: HMSO.

Simons, H. (1995), "The politics and ethics of educational research in England", *British Educational Research Journal*, 21(4): 435-449.