Chapter 5

VALUES, POWER AND STRATEGY IN EVALUATION DESIGN SOME PRELIMINARY CONSIDERATIONS

by

Barry MacDonald
Centre for Applied Research in Education
University of East Anglia
Norwich, United Kingdom

The way we think about evaluation now is not the way we thought about it thirty years ago. That was when, at least in some Western countries, State sponsored curriculum development began to feature as a significant element of post-war socio-economic reform. The major influence on the way we think now about evaluation has been that experience of successive attempts, by various means, to devise and secure improvements in what our school systems provide for children to learn. At the height of the curriculum development movement, in the United Kingdom for example, there were some two hundred sponsored curriculum innovations in action across the country. Today there may be a handful, but to all intents and purposes the curriculum development strategy is dead, abandoned by its political backers as a failure -- it is not effective enough, it is not cheap enough, it is not quick enough. In short, it does not work. What is more, it's all the fault of the professionals, you and me, who had most of the responsibility and therefore must take the blame.

This conclusion is not confined to the United Kingdom. It is shared by the United States, by Canada, by Australia, by New Zealand, and others with a long investment record in the transformational power of voluntary, professionally led initiatives.

And what do we have in place of curriculum development as a strategy for improving schools? We have standardisation of curriculum, we have specification of what has to be done at all levels, we have accountability for delivery of the specification through performance indicators of a quantitative kind. In other words, we have a non-developmental prescription for schooling controlled by output testing. That model is already in place in some countries and appears to be spreading to many others. Within such a model there is little room for the kind of curriculum initiative with which you are currently engaged.

How did we get here, and what are the implications of this massive reverse for the Environment Programme which in some ways, in some of the countries involved, appears to belong to a vanished past? We got here through approaching curriculum development in ways that were based on wrong assumptions about the nature of the problem of curriculum and about the problem of change. As these assumptions were shown by experience to be inadequate or misguided, approaches were modified to take account of experience.

Looking back, we can distinguish a learning curve, or a sequence of phases, each of which constitutes a modification of strategy to take account of previous failures. Now let me say at once that not

every country that has been involved in curriculum development has gone through this sequence in the order which I will describe. Different countries have started at different points in time, and some late starters have taken account of experience elsewhere, as well as of their particular circumstances and traditions. Nevertheless, I would contend that, allowing for variations, approaches to the problem of changing the curriculum have shown an evolutionary pattern; from the simple-minded and mechanistic to the sophisticated and organic. But not enough to save the curriculum development movement.

In the initial phase it was assumed that the problem was simple and the solution could be achieved by simple means. So we commissioned people to write new textbooks and we made them available to the schools. That would do the trick. It did not work. On to phase 2. In this phase, the textbook was replaced by a more imaginative collection of learning materials, and accompanied by extensive guidance for the teacher in the use of materials. This was the packaged curriculum. It did not work either, and by the end of this phase, as the "no significant difference" results began to accumulate, the idea of smuggling into the schools any kind of iron clad, teacher proof innovation had its day. On to phase 3, which involved a switch from teacher proof strategies to teacher led strategies. Here the teacher was seen as the major resource, the motto was "no curriculum development without teacher development", and investment was concentrated on in-service professional development. The idea was that you change the teachers, who will then change the curriculum. This phase was particularly strong and influential in the United Kingdom during the seventies, but it could be said that its very success exposed its limitations. Individual teachers cannot change the institutions in which they work, and the curriculum is institutionally held in place. And so we came to phase 4, where the school is seen to be the unit of change and therefore the target of development strategies.

So we have moved from a textbook phase through a package phase into a teacher development phase and finally school development phase. In that process the people who have led the initiatives have changed their role, from subject expert to teacher developer to organisational theorist.

The main point to note is that throughout all these phases the schools were still failing to deliver the transformations promised by successive cohorts of educational developers. Neither stick nor carrot, nor the various combinations of coercion and incentive that characterised the late period of curriculum intervention, could overcome the apparent intransigence of established practice. For the politicians, who had worked hard on public opinion to land the schools with more than their fair share of responsibility for the economic and social ills of the nation, this posed a dilemma. Ultimately the voters would hold them responsible for doing something about the schools. Their response was to abandon the curriculum development ambition and replace it with a fixed rather than an evolving curriculum brief, and a punitive system of accountability for performance. That is where we are now and that is roughly how we got here.

Presently, I shall focus on the ENSI Initiative, to see what kind of phenomenon it is in this context, but I want first to offer a parallel sketch of how evaluation has changed in response to, and in interaction with, these changes in the approach to the problem of curriculum change.

In the beginning, when curriculum development projects could be defined in very basic treatment/outcome terms, evaluators were attached to projects as measurement specialists, offering scientific proof as to whether or not the student learning objectives of the new curriculum had indeed been attained. Pre- and post-testing of learning gains was the mission of the evaluator. Now that might have been O.K. if the projects had been as successful as they hoped, but they were not and so, whatever difference the projects were making, it didn't show up on the instruments. It appeared that, far from the transformational scenarios confidently predicted by project leaders, nothing much seemed to have happened. Of course, the agencies with major responsibility for master-minding the modernisation of schooling began increasingly to ask "Why?" and they turned to evaluators for answers. But evaluators could not tell them why because they had not been asked to look at why anything happened, just whether it happened.

Within a short time the role of the evaluator changed from that of psychometric technician to that of ethnographer. With the realisation that changing the curriculum was a more complex task than it had first seemed, the task of the evaluator expanded to encompass the interpretation of educational settings, so that the developers and strategists of change might better understand the processes in which they were intervening. And, instead of a narrow focus on student learning, evaluators began to think in terms of assessing the impact of the project, a much broader concept of outcomes. It was during this period that psychometrics gave way to the more naturalistic methodology of case study, portrayal and descriptive and narrative accounts of projects in action. And the notion of authoritative proof of accomplishment gave way to the notion of informed judgement, with evaluators supplying the information.

But still, at this time, evaluators were operating on the assumption that change comes from the top, so that the important audience for their reports were the most powerful actors in the system, those who made policy. It was their understanding, their judgements that mattered. And that meant seeing them as decision-makers. That may sound obvious, but it entailed a big change from the classical position of the detached scientist. Now evaluators had to study the anatomy of decision-making, so that their data matrices might better match the knowledge components of executive choices. Not the rhetoric of decision-making, but the reality. In this way, and not at this time fully consciously, evaluators began to nibble at the exercise of power, and to expand their field of interest to encompass more than the implementation efforts of schools.

As this aspect of evaluation expanded, and as it became increasingly clear that the difficulties of changing the curriculum could not be attributed, at least solely, to the perverse obstinacy of schools threatened by enlightenment, another fundamental assumption bit the dust. That was the assumption that our societies are under rational command, or that at least they would be if only they had the information on which rational command could be exercised. That is not how, in general our societies operate. That finding, in turn made a nonsense of our notion of serving a hierarchically structured decision-making model, and of our simple-minded assumption that what we did as evaluators was, in political terms, non-problematic.

Suddenly, it seems now in retrospect, evaluation became politicised in its consciousness, and highly problematic as an occupation. New questions, hitherto not posed, came to the fore. "Who am I helping with what I am doing?" became a central issue, an issue not unrelated to the problem of change as we identified blockages to change far removed from the discretion of schools, teachers and learners. We became aware that we were part of a process of strengthening the knowledge base of the more powerful actors in the system at the expense of the less powerful, and moreover without securing a noticeable benefit to educational development in return.

For some of us, that meant another shift of role, from the managerial feedback role to what we might call a mediation model of knowledge creation and transfer, a negotiation model, or if you like, a broker model of evaluation, in which the interests, rights and obligations of a whole range of constituencies have to be taken into account in deciding what knowledge is to be generated, for whom and about what. For years, we had been happily extracting information and knowledge from the school system, without putting much of its back, on the assumption that such data would be used to benefit schooling, and for no other purposes. That will not do any more. Nothing in evaluation is straightforward. Everything we do has to be justified, not just technically, but ethically and politically. It is not just curriculum developers, but evaluators too, who operate, implicitly or explicitly, with a theory of change.

You can see in this account some parallels between the thinking of developers and the thinking of evaluators. The experience of developers led them to reject the notion of securing change in schools through packages devised outside schools and sent back in for teachers to implement. In place of that they sought more investment in the grassroots, in the human resources of the schools, in the notion of schools as sites of continuous auto-regeneration, whose diverse creativity would shape and guide national policy of a responsive and supportive kind. Action research evolved as a major expression of that strategy of

change, and of the politics of teacher professionalism. In like fashion evaluators, looking to make a more influential contribution to development, are less inclined to prioritise management needs, more inclined to prioritise direct feedback to the workforce, and keen to explore collaborative processes of evaluation that can help to internalise learning.

Now I think that through all these cumulative changes we can see a clear line of development with regard to the definition of the evaluation task and the concept of the role. Beginning with the conventional research paradigms from which evaluators were recruited, there was the tendency to see programmes as quasi-experiments from which, concentrating on their amenable aspects, some scientific generalisations could be usefully extracted. This phase was followed by a more holistic view of the innovation as an intrusion into a settled culture, which had to be interpreted and understood if it was to be effectively influenced. That was the ethnographic or anthropological phase, utilising an expanded database and increasingly concerned with detailed study of the contexts and processes of new implants in the school system. This was still largely focused upon the experience of implementation at the grassroots level. Case study at this point meant case study of innovating schools. Then came, as a result of continually expanding the framework of explanation of school-based phenomena, a redefinition of innovation as an intrusion into established systems of power, and as attempts to redistribute power in support of new values.

At this point the concept of the case was redefined as the whole programme, including participants at all levels of responsibility for its origination, management and implementation. It is not hard to see the political implications of such a redefinition, or to understand how that changed the role of the evaluator to that of someone who is mediating power relationships by virtue of choices she/he makes about who gets to know what about whom. And if you see, as we do in the Centre for Applied Research in Education, the public function of evaluation as holding the exercise of delegated power to civic account, with respect to all levels of action, then this puts a premium on the independence of the evaluation.

I imagine that most of the participants at the Cromer Conference, as Pedagogical or Administrative Support persons within the Initiative, see their role in some sense as a mediating role within hierarchical power systems, both protecting and stimulating activity in the schools and districts and at the same time trying to satisfy those "up there" who are financing the initiative. All I am saying at this point is that evaluators also have a mediating role.

So the programme is the case, and case study evaluation is a study of what particular people in the case do and think in their specific circumstances at particular points in time. Now let us ask, with respect to the Environmental Education Initiative, what is it a case of? I will tell you what I think it is a case of. I can do that because I come from outside it, I have no involvement with it, and I am not going to evaluate it. So here are some thoughts, based on the little information I have gleaned from conference papers. Clearly, it is an interesting case. At one level, it is a wholesome and harmless initiative, which has little children going out and planting perhaps a beetroot? What could be more wholesome than that! At the other level, it is an international conspiracy. Is it a specific curriculum innovation, or is it an attempt to take one single issue and restructure educational systems through it? It is fairly clear that some people had that in mind. Is it, in fact, an attempt to get an international response to a problem which can only be solved internationally? Here we have a planet in serious danger. We live increasingly in a habitat that nobody wanted and, in a sense, nobody made, but which we do not seem able to do anything about. So, at one level, this initiative can be seen as the first stage in a grand plan to put a brake on that process and, if possible, to reverse it.

Now, this is highly relevant to how we plan evaluation. Some may say, 'the answer has to be international in order to get any leverage in our own country'. In this case the international aspect becomes important. So that, in conducting an evaluation those with that view, have to take seriously what use can be made of the data in Paris if in fact that level of action or influence is important to the overall plan. But there are other things to take on board. Many of the ENSI participants were involved in this initiative before evaluation came along. Some may have been commanded to engage with the initiative, some may

have had nothing better to do, but most have probably a strong commitment to it. If this is true it raises the question, 'what are you going to do about evaluation?' If I come in, I have no such commitment. I can come in and look at this initiative from whatever perspective seems to illuminate it and make it understandable to people outside. But I would say to participants, 'if it is a conspiracy to which you are party, then you might want to look at it quite differently. You might want to say, "what can we afford to do strategically at this point in time?' "Now one of the things that is interesting about the papers I have looked at is the delicate issue of this being seen as a political initiative, as a dangerous form of subversion. And I notice that reassurance is being given to those who might harbour such a deeply wounding suspicion. Children have to be introduced to the realities of the world, to the costs and the benefits, to the fact that nations are all having to compete economically to survive, that industry is essential, various forms of pollution have a certain justification. But nevertheless underneath that there is a view that if this kind of initiative gets looked at critically in that kind of light, too soon, it will not yet have gained enough purchase to have a future. Now, obviously, strategically one gets an initiative like this by going in the "soft way". It is another harmless OECD project for volunteers; those naive people in the schools, who happily do a little bit on the environment and keep everybody happy by saying, 'yes, we are doing something about environmental education, yes' -- and so forth. But, we know enough about the history of innovations, we know that so long as it is at the margin it will get assimilated, ignored or shut out of the school. Either way, it is no good. It has got to be moved into the centre of the curriculum, into the mainstream.

My question to ENSI therefore is, 'what is your strategy and timescale for doing that? Does your investment still at this stage need to be on improving the quality of the work in the schools, or is the emphasis to be shifted to another level? Do you need to do anything about the networks you have got, to strengthen those? Is that where the action is in the next phase?' It is very interesting because here is an initiative, a curriculum development initiative, which has already been going for seven years and there is no reason why it cannot go on and on. Now that is unusual because nearly all curriculum developers have been asked to transform the world in less than five years, sometimes three. Now here an interesting issue arises. Is this a case of an educational initiative, or is this a case of a movement, which is a different thing? Is it a movement, or is it a finite curriculum development? What is the case? What is the strategy of change? There are two tasks: one is to realise the values in a viable operational form in schools -- that is to specify the curriculum in action; and the other one is to secure it -- to root it in the system so that it can grow. Both these aims have to be part of this strategic thinking.

Now, those who are outside evaluators, like me, would just go in and would report everything that is happening, all aspects of it. But participants are mostly inside, probably with some kind of commitment to the enterprise. So for those people, what does evaluation mean? What kind of evaluation? What is it for? Is it an insider evaluation to strengthen the inside, in other words to supplement the product of the action research, which already generates a large amount of formative data? Is that what it is? Or is it insider evaluation for outsiders, to influence them, and, if so, have participants got to be selective in how they represent the work, or not selective? Or are some of them outsiders looking in? You can have an outsider evaluation for insiders, or you can have an outsider evaluation for outsiders. I would say to participants, 'What you do is a question of where you stand in relation to the future of the Initiative and the contribution your evaluation can make to that future.

There is danger in evaluation and there is promise. And there are always unanticipated consequences, although sometimes these are confused by unacknowledged intentions. In designing evaluations, try to be clear about your values and priorities in relation to the values and priorities of the Programme. And about your evaluation strategy. There is high risk and low risk evaluation. A high risk evaluation, which is honest, which reports everything as objectively as possible, may gain credibility for the initiative, or may expose it, perhaps prematurely, to unfair criticism. A low risk evaluation, cosmetic or compromising, may do little harm, but may fail to stimulate the kind of constructive critique that can help the Programme to improve and help others to understand the levels of support needed if the task is to be accomplished.'

The question of priorities is an important one for a group like this to consider. At this stage, for instance, how important is the international nature of this initiative? If it is important, then participants need to pay attention to the extent of their individual contribution to the construction of an overview. Comparable data, common foci and themes will feature prominently in such designs. On the other hand for some, the priority may be influencing own state national policy-making, or developments in a particular locality. It may be for others, that interest in this initiative is not so much an interest in environmental education *per se*, but in exploring the feasibility and fruitfulness of classroom action research, or in the possibilities more generally of ways of improving schools. In that case, it could just as well be a programme in drugs education, or information technology. When I read ENSI documents I see an emphasis on breathing new educational life into atrophied school systems, and I do not care which country we are talking about, that need will be there. This initiative has a very strong educational line, and one that is readily generalisable to the whole process of schooling. This is high ambition, and we know from thirty years of evaluating innovations in schooling that it is the truly educational component that is most resisted, the educational values that get lost as innovations are stripped down for assimilation into the mainstream curriculum culture.

Let me conclude by offering this advice to participants: it is important to be clear about what you are up to in this particular innovation. What is your view of your school system, its functions and its possibilities? Schooling can be seen as a conspiracy by the State against people, or as a conspiracy by educators like us against the State, or as any number of things in between. Where do you stand? All I am saying is that at the outset, when we are going to design evaluations, issues of this kind cannot safely be left off the agenda.

What are you going to do and why are you going to do it?

What is the Initiative a case of?

Do you want an evaluation to help the internal development of the Initiative, or one which is about informing outsiders, either because they are entitled to know and to judge for themselves, or because you feel that unless you begin to do that the Initiative will remain powerless?