

3e ✓

EDUCATIONAL EVALUATION IN THE CONTEMPORARY WORLD\*

Barry MacDonald  
Centre for Applied Research in Education  
University of East Anglia  
England

I have been told that this conference is an innovation, something new in Spain. Being an evaluator this interests me. Innovations are my business. Almost all professional evaluators are employed to study innovations. Educational innovation is a big industry in the Western hemisphere, and it is coming your way, I suspect.

People in Spain clearly share that expectation; that is one reason for this conference, one reason why some of us are here. And when it comes, this innovation industry, when the pesetas are made available to initiate programmes of research and development concerned with curriculum reform, to bring schools into line with the needs of contemporary Spain and contemporary Spaniards, then you will generate a new specialist form of investigation - called evaluation.

We evaluators are the camp-followers of education reform movements.

Now why is it that only the new needs to be evaluated? Why not the old, the established, the familiar? Would it not make more sense to evaluate the present before launching into the future.

There are many answers to this question:

1. There is no time - the need for change is urgent.
2. There is no need - we know what is wrong.
  - a) curriculum is out of date
  - b) we need more engineers
  - c) the Russians are ahead of us
  - d) too many illiterates

\* Invited presentation to the *Simposium Internacional De Didactica General y Didacticas Especiales, La Manga del Mar Menor, Spain, 27 September - 2 October 1982.*

What we do not know is how to put it right. Innovators of course claim to know how, and in 30 years none of them have admitted to being wrong though their evaluators maintain they have all failed.

3. Innovation costs money, and has to be justified. Of course established practice costs money, but the need for that expenditure is not contested.
4. Innovation can get out of control, not just financial control but ideological control.

We evaluators are the policemen and accountants of educational reform.

We make sure that nobody spends the money on holidays in LA MANGA. It's okay, I'm not here to evaluate the conference. But suppose I were? What if I were here to evaluate this particular conference? What would the problem of evaluation look like?

Of course it would hardly be right for me to do so - after all I am part of the conference - I have an interest in its success. It would be reasonable to assume that my evaluation might be contaminated by this interest, this responsibility and involvement. I lack the necessary disinterest that evaluators must have in the innovations they study.

But I could of course evaluate my own performance today as a separate entity. This would not free me from a sense of obligation to my hosts with respect to the rest of the conference, but nevertheless it would reduce the level of contamination. In any case, wouldn't you expect a professional evaluator to evaluate his own performance?

Well, I'll let you into a little secret. I have made all the arrangements for an exemplary evaluation of this little event, illustrating the newest techniques. You see those cylinders fixed to the ceiling, you think are light, don't you? Actually, they are the latest invention of the Californian school of quantitative evaluators - they are meters for measuring the frequency of your eye-blink during my presentation. This is an impact measure. It is negatively correlated. High frequency means a low impact - low frequency means high attention. Of course, as with all instruments of evaluation, this one is still at an early stage of develop-

ment. For instance high frequency could indicate an outbreak of conjunctivitis, a very low frequency a state of catatonic trance. That's the trouble with these quantitative instruments - we know they measure something, but we are not sure what. And there is another important point about such measures - they are easy to frustrate. If you had known I was planning to evaluate this meeting, and if you had also known about the blink-meters, you might have come in dark glasses. Then the machine can't tell even whether you are awake or asleep.

But why would you want to play such a dirty trick on me? After all, it is my performance, my curriculum, that is being evaluated. The answer is that you cannot be sure about that. The data I need to evaluate my performance is data that could be used to evaluate your performance. So before you decide whether to wear dark glasses or not, you need to ask some questions:

- Who is this evaluator?
- Who is paying his salary?
- Whose questions are being answered by this evaluation? - Because you know that those who control the questions control the political utility of the information.
- Who has access to the evaluation report?
- To what uses might the data of this evaluation be put?

We have to take account of the fact that most evaluators are paid by employers to look at the work of employees. After all, evaluation costs money, so it is those who control resources who fund evaluation. And we have to note that the usual result of evaluation is a redistribution of information about educational practice in favour of the more powerful. And information is power.

Now I could say - I'm not doing any of those nasty things - Goodness me, I have to sleep at nights. No, my purpose is to understand the deficiencies in what is provided for you, so that next time the curriculum that you get will be more effective. My purpose is to help to improve the service. As the Conference Questionnaire says: "Participar es majorar". Hotels have the same rhetoric. Perhaps at the end of this week we will have a little questionnaire from our hotel manager. It will ask - do you like your eggs

boiled for 3 minutes or 6 minutes? Very nice - they want to improve the service. Only, if enough of you say 3 minutes, then next time there will only be 3 minute-eggs. Nobody asks, "How do you like your eggs?" But then, that is clearly a managerial evaluation. Alas, we get those in education too. One of the biggest, most expensive evaluations to take place since the Second World War was funded by the United States government. The evaluators were asked to look at 20 different innovatory curriculum programmes, each offering a curriculum for low-achieving children in the elementary schools. These curricula were all very different - in their values, aims, content, methods, populations and locations. But the government wanted to know which of them were the best - they wanted the best two out of the twenty, so that government money could be confined to supporting those two. They wanted the options reduced to 3-minute-eggs and 6-minute-eggs. And the evaluators did it. It took ten years and 20 million dollars to find out, but in the end they put the programmes in a rank order. There was uproar, so much so that the government was unable at that time to fulfil its intention.

You see, what the evaluators did in the end to achieve this great feat was to ignore the variety of aims and circumstances, and instead to judge the merits of the programmes largely by administering a few simple tests of achievement to all the children involved. These tests were inevitably biased in favour of those programmes whose curriculum happened to emphasise the same achievements as the test developers. Now the evaluators did not want to do this - it was not their intention when they started out - they had reassured all the programme developers that the evaluation would be sensitive and fair to their individual differences. But they came under increasing pressure from the central bureaucracy to come up with the required result - two clear-cut winners.

But why did they do so? They did so partly because they belonged to a commercial business of policy evaluators, dependent for their living on government contracts (in the past 20 years many such businesses have sprung up, especially in the U.S.A. but recently in Europe as well, and they compete very successfully with the academic community for government money). They did so partly because they believe that their function is to help those who employ them, to legitimate their position of decision-makers, and to enable the managers of the system to solve management

problems. Evaluation is after all a utilitarian mission - its function is to help people to take the next step, not just to become wiser about the social realities they are responsible for. The problem for the United States educational administration in this instance was too much uncontrolled variation in the system, too many programmes, too many competing claims for federal support. It was the task of the evaluators to reduce all this noise in the system to a simple set of options, each of them assessed in terms of cost-benefit. Evaluation was seen as a key element in an effort to impose a rational decision-making model on a very complex social system. I am here talking about evaluation in the context of decentralised educational systems such as obtain in the United States and in England. I am not talking about Spain, or France for that matter.

One has to understand that the political and economic context of education in many western countries has become transformed in twenty years, and this has affected the context and the practice of evaluation. Let me dwell on this transformation a little, because it has significant implications for Spain. I am here contrasting the circumstances of the sixties, when educational evaluation rose to prominence, with the circumstances of today.

In the sixties evaluation was an instrument of growth and investment in buoyant economies and expanding educational provision. It was concentrated on innovations, helping to assess the discrepancy between hopes and achievements so that more resources could be effectively directed to meet new and challenging goals. It was part of the curriculum development movement first in the USA, and then in Britain and other European countries, through which a systematic attempt was made to bring the curriculum up to date, at least as up to date as the Russians and the Japanese. That curriculum development movement was itself part of an economic movement to acquire a competitive edge in the post-war international market. Education was seen as a key to national success, innovation was the prime need, and evaluation provided the quality control - by checking out the claims of the innovators, identifying new needs, indicating new possibilities for problem solving. That was the context of evaluation in the sixties.

Such a view of evaluation would be hard to sustain now. Galloping recession, a context of cuts and resource starvation. Investment has dried up or is confined to essential system maintenance. Curriculum development on the scale and range we knew in the post-war boom has run its course, having apparently failed to deliver industrial solvency. Its main agencies in England, the Nuffield Foundation and the Schools Council, have declined in power and resources, and Government funding is confined to a narrow and tightly defined set of priorities under executive control. Evaluation has continued to thrive, but more and more often in the role of an instrument of accountability, of good housekeeping, of fiscal and managerial control. Whereas the funding of an evaluation used to represent a commitment to whatever educational enterprise it was attached, now it is just as likely to represent hostility or mistrust. When the British government set up an evaluation of the Schools Council it had already decided to close it down and was looking to the evaluation to provide the ammunition. This is akin to the trial in Alice in Wonderland where you start with the execution and proceed to the evidence. Nowadays if anyone suggests to you an evaluation of your work, you'd better check to see if they are interested in developing it or dispensing with it. That's why its getting much more difficult for people like myself to negotiate access to educational institutions and to gain the trust and co-operation of people in them.

TWO other important changes have taken place in educational evaluation over the period of time I have been considering.

1. Firstly in the amount of evaluation - there is more evaluation, particularly more policy evaluation, less research, especially basic research.
2. Secondly there is a shift in control of investigation away from the universities and nearer to the centres of political power, especially bureaucratic power. Government departments are now the main sponsors of evaluation studies (the role of the charities has declined) so there is less independence for the evaluator in designing or conducting or disseminating their work. It means greater docility on the part of people like me to a monolithic control of the processes of information production.

Associated with this is a change in the marketability of different models of evaluation. Back in the early sixties, the model of evaluation then prevalent was associated with behavioural objectives. You took any educational innovation, turned its aims into learning objectives, and employed evaluators as measurers of intended outcomes. In the course of the sixties, this model came in for a lot of criticism. It was not informative enough, not helpful enough with explanations of shortfall so there was a big movement, of which I was part, away from the measurement of outcomes towards more descriptive/interpretive studies that concentrated on depicting the circumstances processes and difficulties of innovation, offering a better understanding of the problems of introducing new practices. Naturalistic forms of enquiry evolved in this context.

This alternative movement in evaluation was making a lot of headway and by the seventies had become the dominant approach in England to the evaluation of educational activities. Then came the recession, the rise of managerialism, the shift in control I have talked about, and increased pressure from the sponsors of evaluation for hard, simple, quantifiable measures of success and failure. This is perfectly understandable - with the shift to policy concerns and centralised decision-making, we found ourselves increasingly dealing with decision-makers with neither the time nor the patience to deal with complex information, nor with the contradictions or ambivalence that tend to characterise particular cases of educational action.

This kind of evaluation tends also to pose another difficulty. When as an evaluator you look very closely at what educators are doing, and you are able to take full account of the constraints under which they operate, you may well conclude that those activities need changing if they are to become effective, but you are also likely to vindicate the actions of those responsible as being intelligent people working with integrity on difficult problems. The implication is usually that these people need more help with the problem and this is not a conclusion that universally finds favour with those whose priority may be to cut out wasteful activity, and to include the people involved with the waste. For this purpose an aims/achievement model of evaluation affords far more freedom, since it will inevitably reveal failure without explanation, making it much easier to blame the performers rather than the providers of the service.

The point about the simplicity or complexity of evaluation designs is crucial to the question of who controls education, and its the main point I want to make. Let me give another illustration, this time from my own experience.

During the seventies, in England, I was responsible for the evaluation of the single, biggest curriculum development project that England had seen since the beginning of the curriculum development movement. The project was concerned to promote computer-based learning, it lasted five years, and it supported about 40 different experiments throughout the United Kingdom. It cost 2½ million pounds. It was the first such project to be administered and controlled directly by government civil servants - it was run by an executive committee of bureaucrats from seven government departments, all with some direct or indirect interest in the welfare of Britain's computer industry.

This committee made all the decisions about whether to fund a particular experiment, whether to continue to fund it or to terminate its support. They met 5 or 6 times a year to make such decisions - and to help them they employed a team of evaluators, led by me. We made progress reports to the committee on how the various projects were getting along.

We quickly ran into some of the problems you might expect from what I have been saying. I will try to summarise the main issues.

1. They wanted an aims/achievement model of evaluation. The funded projects were required to state their objectives before they got any money. They wanted an evaluation focus on whether those objectives were achieved.

We said no - that was unfair. None of the projects would achieve all their objectives - so they would all fail on that criterion. We would portray their efforts to achieve those objectives so that the Committee could judge whether they were engaged in worthwhile activities, given their constraints and opportunities.

2. They wanted us to make recommendations about which projects should be supported, which terminated.

We said no - it was their task and responsibility to make such judgements, not ours. They would have to read the evaluation reports, and make up their minds.

3. They wanted us to add to the written reports - to tell them things about the projects we could not put in the reports.

We said no - no secret reporting. Our reports were all negotiated with the people whose work they commented on, and not given to the Committee until those people agreed they were accurate, relevant and fair. We would not add to them.

4. They said the reports were too long and too complex for a busy committee to deal with. Could we not summarise them?

We said they were as short as we could make them and still negotiate them with the people whose work was being judged.

They said - but we cannot handle all this complexity - are you saying we should not be making these decisions?

We said - that is for you to say, and for others who read the evaluation report on the work of this Committee.

5. They said - what do you mean, the evaluation of this Committee? We don't want you to evaluate us, only the projects.

We said - we're afraid we must - it would not be fair only to gather information about the projects. They want to know about how you do your work, whether you reached your objectives, whether you are doing a good job.

6. They said - who do you think you are? We are paying you to do as we say.

We said - we are your democratic independent evaluators. Simply because you pay for the evaluation does not mean that you have any special claim on its services, or exemption from its focus. You cannot buy an evaluation, you can only sponsor one. Anyone has a right to raise questions and issues for inclusion on the agenda of the

evaluation, and no-one has the right to ask for information without being prepared to give it.

These arguments with the bureaucracy were bitter and prolonged. They went on for five years. I raise them here only to indicate the lines of a counter-argument to bureaucratic assumptions based on power and custom rather than responsibility and reasonableness.

My point is that evaluation is in a significant sense a political activity and a valuable political resource. I've already suggested that as such it has become increasingly controlled by powerful agencies who try to impose upon its design and methods the kind of structure that is consistent with the maintenance, extension and legitimation of that power. Now this situation, in so far as I have accurately depicted it, offers you three possible responses.-

1. Don't have anything to do with it and hope the contagion doesn't spread into Spain.
2. If it does descend upon you, put on your dark glasses, and subvert it by any means at your disposal.
3. Take the initiative, develop your own evaluation, your own political resource that you can utilise in pursuing your own aims or in opposing evaluation-based threats to your programme.

Now it may seem to you, in view of all I have said, that this third option is a very poor prospect. Let me argue against that by saying that evaluation can be the instrument of decentralisation in situations that suffer from too much central control. Up to recently, nearly all evaluation was sponsored and carried out by people outside the institutions and activities being evaluated. In recent years, a movement towards self evaluation at an individual and institutional level has emerged. A growing number of educational institutions and individuals are involved in some form of self-monitoring. This is:

1. Partly a response to demands for accountability - the pressure on educators to give accounts of their work - to show that they are gainfully and effectively employed.
2. Partly a response to the appeal of evaluation as a means of

improving what educators do (we might call this the naive view of evaluation, the kind of view that we started with in evaluation but which has since become more problematic).

3. But partly it's a defence/offence in response to the kind of dangerous and unfair external evaluation practices that have been growing rather popular with educational administrations
- 

Those who initiate evaluation studies of their own work are seeking to exploit the advantages of a proactive policy. Self-evaluation provides opportunities to

1. Describe fully the nature, values and effects of what they are trying to do. Outsiders tend to understate the purposes of educators and the impact of their efforts.
2. Articulate the questions that it would be reasonable to seek answers to.
3. Propose the criteria by which their work could fairly be judged.
4. Generate their own ammunition in the contest for shares in the shrinking resource cake, or at the very least a basis of defending themselves against cuts.

It is important that the evaluation capacity is decentralised, because then it can legitimate educational diversity, whether that diversity is the result of professional conviction or circumstance, and mount a challenge to decision-making power structures that inhibit healthy growth and experiment in the educational system. Unless you, at various levels in the system, engage in evaluation, then the evaluation you get will merely consolidate the power structure you already have. This is the political significance of the work that Professor Elliott described the other day - it is a means of strengthening the critical and defence mechanisms of a diverse educational community.

So, don't put on your dark glasses yet. Rather be prepared, when evaluation comes along, to press the critical questions: Who does it? Whose questions get answered? To whom does the data belong?

I wish you luck with evaluation in Spain.