Introduction

The idea of ‘evidence based health care’, when it first came to be articulated, and indeed became the basis of a Masters course at the University of Oxford, struck me as odd to say the least. It came as a surprise to learn that decisions over the most appropriate form of medical treatment were not necessarily based on evidence – for that seemed to be the implication. Medical prognosis and treatment, I had taken for granted, would be based on scientific research. Indeed, to qualify as a doctor required a rigorous background in the relevant science.

However, the Cochrane Centre in Oxford, under the leadership of Sir Ian Chalmers, did not have such confidence. Much practice was, as it were, ‘inherited wisdom’. Other practice, apparently based on evidence, competed with other practices based on other evidence. Any one who has sought medical help for back problems will know what I mean.

In many respects, this should not be surprising. A merely superficial acquaintance with the philosophy of science makes it clear that science grows from constant (and often successful) attempts to negate the current state of scientific knowledge. All such knowledge is, as it were, provisional – to be accepted until such time as it is refuted and replaced by more comprehensive and better corroborated scientific propositions. But such development in science, such gradual ‘approximation to the truth’, depends on the application of rigorous scientific methods whereby error is to be eliminated. The Cochrane Centre believed that such rigorous methods were often lacking in the arrival at medical ‘truths’ which informed and shaped medical practice.

Such methods required strictly controlled experiments, with very large control and experimental groups, so that one might see clearly the difference which a particular intervention might make. But this in turn required extremely careful articulation of the
hypotheses to be tested, and that in turn required a sophisticated process of refining the felt problem (felt as often as not by the patients as much as by the medical researchers) into a testable set of hypotheses. Furthermore, such large scale experiments would have to take into account the research which others had conducted in similar areas. But that required putting together pieces of research which, more often than not, were based on different samples or which made slightly different assumptions. Hence, an important part of refining the evidence lay in the systematic review of existing research, rejecting that which did not meet rigorous experimental criteria, ignoring that where the data and method were less than clear, reconciling where possible the different bases for the samples, identifying where further research was needed to fill the gaps in our scientifically based knowledge. These systematic reviews and the subsequent meta-analyses of available research (which are described by Davies, Gough and Andrews in this volume) were difficult and time consuming. They required co-operation across continents. They required an explicitness of, and openness to, the problem, the hypotheses, the sampling process and the data. They encouraged open and critical debate, and the constant refinement of the conclusions in the light of that critical debate and new data. Indeed, ‘evidence based’, so conceived, had a thoroughly Popperian ring to it.

Such has been the success of the Cochrane Centre’s work that people in other areas of the public services have looked for the lessons which can be learnt from it. The Campbell Collaboration, based in the United States, but with regional centres in Canada and Denmark, has extended the work to other areas of social life – for example, education and criminology. This is described very well by Davies in this volume. Both the Secretary of State for Education and the Home Office saw the approach of Cochrane and Campbell to be what was required to improve the quality of research to inform both government policy and professional practice. And this was seen to be necessary because of the criticisms of that research, certainly in education. The criticisms were several, but they might be summarised as saying that the research was: too fragmented (too little of the large scale and ‘bold’ hypotheses thoroughly tested); based on different assumptions, samples and data; often less than rigorous in method; not unambiguously address to a specific question to which the policy-maker or the practitioner needs an answer.

The reaction to the transfer, to the field of education, of the evidence-based approach of Cochrane and Campbell has varied from the hostile to the welcoming. That is reflected in this volume. But the essence of these criticisms and of the differences between them is philosophical. It concerns the nature of research, and that in turn the nature of knowledge. What counts as evidence for a particular kind of knowledge claim? In this
paper I shall briefly outline what I think are the key philosophical difficulties, not simply in
the adoption of evidence-based policy and practice, but also in some of the criticisms if it.

These philosophical issues are: first, the nature of ‘evidence’; second, the extension of
the methods of the natural sciences to understanding of human beings; third, the
adoption of a means / end model of educational planning and decision making.

Evidence

A lot depends on how one interprets the word ‘evidence’ There are many different kinds
of evidence, depending on the type of claim being made. Evidence that water boils at 100
degrees centigrade at sea level would be very different from the evidence to indicate that
a rock face is 100 million years old or that Caesar really did cross the Rubicon or that
Saddam Hussein’s regime was evil or that Freddie has been a good boy. There are
different forms of discourse, each characterised by different ways of looking at the world,
different kinds of truth claim, different ways of investigating the truth. What counts as
evidence will depend upon the kind of discourse one is engaged in. Historical evidence is
different from that in science, and even within science there are different sorts of
discourse, each characterised by differences in what is deemed to constitute evidence.
Hence, there is a danger of criticising a piece of evidence because it does not meet the
standards of evidence in a quite different form of discourse. Indeed, that is the cause of
certain problems within the arguments ‘for’ and ‘against’ evidence-based policy and
practice within education. Some, who advocate ‘evidence-based’, do so by blurring the
boundaries between scientific and non-scientific forms of discourse, thereby rejecting
certain claims as without foundation. On the other hand, certain critics, by identifying
evidence-base with only one sort of evidence, reject entirely the idea of ‘evidence-based’
as irrelevant to the complex problems of educational policy and professional practice.
Furthermore, ‘evidence’ must not be confused with proof. I have evidence that John told a
lie, but I cannot prove it. One can gradually build up the evidence for a belief but
gradually proving it seems a little odd. On the basis of evidence, it may be probable
something is the case – although there may be counter-evidence which is less
persuasive.

These comments are by no means irrelevant to the educational debate. Often politicians
seem to advocate an evidence-based policy as though one should only act when one can
demonstrate that a particular course of action is proven to be the correct one. They feel
let down when the research, on the basis of which a particular policy is adopted, turns out
to be less than adequate. But all one can say, as a result of research, is that in the light of
all the evidence, and balancing the evidence both ‘for’ and ‘against’, one course of action seems to be the most rational one to adopt. And, indeed, that may well be the case, until such time as contrary evidence is discovered. Furthermore, the evidence upon which one acts can be weak or strong, and, very often, one has no alternative but to act on weak evidence. The teacher, faced with a quick decision over the treatment of an offender, has no time to find conclusive evidence. ‘Deliberation’, followed by ‘judgement’, requires a quick survey of different kinds of often weak evidence before action is swiftly taken – proximity to the scene, previous record of similar behaviour, a prima facie motive. Indeed, evidence here is much more like the notion of evidence in a detective novel than it is in scientific research. And notions of ‘deliberation’ and ‘practical judgement’ (which goes beyond the available evidence) cannot be avoided.

Furthermore, educational discourse is eclectic. It draws upon different kinds of evidence – scientific certainly, but also personal insight, historical, psychological. What is to count as evidence in any one situation will depend on the particular educational judgements being made, and generalisations will always be negated by particular cases – a point which I shall develop in the next section. Thus, educational practice requires judgements about ‘achievement’ as well as about the ‘ability to achieve’ and about the ‘capacity to have the ability to achieve’. It requires judgement about intention as well as motivation. But no amount of observed behaviour, though logically related in normal circumstances to having certain intentions (for example, to complete one’s homework) or to being motivated in a particular way (for example, to please one’s parents), proves or means that such is the intention or the motive for action. There is always a logical gap between the conclusion and the evidence for the conclusion.

These preliminary remarks on the concept of evidence are an introduction to the more philosophical issues arising out of the papers in this volume. For at the heart of the understandings of evidence-based policy and practice, and indeed of the arguments about the importance we should attach to it, are philosophical issues about the nature of evidence, of proof, of knowledge within the social sciences and of educational discourse and judgement.

**Philosophical issues**

I want to pick out three interconnected philosophical issues which, in one way or another, arise in the various contributions. These are:

- first, the logical unpredictability of all the consequences of a particular course of action or a particular policy;
second, the irreconcilability of scientific discourse (and thus the social sciences within a particular tradition) with that concerned with persons; third, the logical separation of educational ‘ends’ or ‘goals’ from the ‘means’ of achieving them.

**Unpredictability**

The first issue concerns the difficulty in predicting what will happen if … in complex social situations. (It is an argument developed very effectively by Luntley, 2000, in connection with the proposal for performance related pay.) Thus, the government, in the light of evidence, believes that a particular policy will have certain predictable results. And, indeed, from the government’s point of view, research should be indicating what consequences will follow from certain policies. What practices are most effective in achieving the desired results? However, there are two senses in which this cannot be the case.

The first sense is that the agents, whose actions are being predicted by the adoption of such a policy, change the context in which the predictions are made once they are aware of what is being intended. Once the pupils are aware of the rationale for the emphasis upon literacy strategies (for example, raising the scores of the school and thus the position of the school in the league table) so they are able, and might be willing, to subvert the policy. Such changed consciousness and its effects could not themselves be anticipated in the development of the strategy, however evidence-based that was in the light of previous practice.

The second, and connected sense, is this. According to Luntley (ibid. p.17), classrooms (and other educational units) share a common structural feature with other social and natural systems – namely, non-linearity. Ignore this and you get a faulty logic of understanding of the system at issue.

Within very complex systems of interacting elements, especially when those elements are endowed with intelligence and where the interactions are consciously engaged in, the full impact of all these millions of interactions cannot be predicted with accuracy. And the impossibility of so doing is not just a matter of size and complexity. Rather (Neither???) is it (???) a logical matter, for one interaction changes the nature of the situation such that the effect of x upon y will not be the same the second time round. The many different elements in the situation are interacting with each other in a way that cannot be controlled
from the centre, and they are thus changing the context which the centre wishes to control and influence. In economics, the countless interactions in the market place constantly change the context in which macro-economic management is meant to take place. The tax changes also ‘tax’ the imagination of those who seek new and unpredicted ways of dodging the taxes, thereby creating a different economic and social situation, which in turn makes unpredictable demands upon the economy and so on.

Given this necessary unpredictability of complex social situations, there is a limit to how far the accumulation of evidence can ensure certain consequences will follow from carefully considered interventions.

**Explaining human behaviour**

Educational policy and professional practice are ultimately about getting people (usually young people) to learn something – and something which is deemed to be of value. To educate is to develop the capacity to think, to value, to understand, to reason, to appreciate. These are states of mind, mental capacities, distinctively human qualities. One feature of such states of mind is that they constitute a different kind of ‘reality’ from that which is the subject matter of the natural sciences. I can observe tables and chairs; I cannot observe in the same way intentions, motives and thoughts. The hand raised is seen by everyone, but may well be interpreted very differently – a wave to a friend, a request for attention, the signalling of a revolution, an expression of exasperation. The understanding of that behaviour depends on knowing the intention – and the motivation for so intending. Thus, the request for attention (the intention of raising the arm) could be motivated by boredom or by excitement at a discovery or by the wish to annoy. Explaining human actions requires reference to intentions and motives, not to causes (generally speaking).

Furthermore, those intentions and motives presuppose a social context of rules whereby the intended behaviours are going to be interpreted by others in a particular way. It is no good signalling a revolution if the fellow revolutionaries do not understand the gesture. To explain human actions requires a grasp of the social rules through which social intercourse is able to take place. Furthermore, such social rules will change from social group to social group – indeed, a social group is partly defined in terms of the social rules through which they engage with each other. There is a set of expectations amongst allotment owners which shapes their behaviours in a way which would not be fully understood by those who have never been apprenticed to this form of life. And no doubt these rules and expectations vary from one allotment to another as populations and
economic circumstances change. (Candlelight dinner parties are held by some allotment holders on our patch, but none the less within a social context which inherits certain expectations from previous allotment holders). Explanation, without reference to such social rules and context and without recognition of their variability according to different social and economic circumstances, is not an explanation of the human world we inhabit.

Certain consequences about evidence-based policy and practice are drawn from these considerations – some valid and some not so. First, the distinctive nature of human explanation must set logical limits to large scale explanations of behaviour, whether educational or not. Such large scale explanations cannot be sensitive to the complexity and variability of social rules and expectations through which decisions and actions are made intelligible. The significance of being numerate or literate, the value of higher education, the respect for the teacher, an interest in literature, and so on will be different from one social group to another – whether such groupings are determined by ethnicity, religious tradition, economic affiliation, social class, regional history or family allegiance. What might ‘work’ in one context might not do so in another, and the reason might be partly explicable in terms of the social rules and the institutional framework (of family, of religious faith, of civic custom) within which the agents are making sense of the world, finding value in some activities rather than others or developing relationships of a particular kind. That is why evidence-based practice needs to look carefully at the particular contexts (the implicit rules and expectations which shape behaviour and which are sometimes embodied within the institutions the learners belong to) in which professional judgement and decisions are to be made.

However, a second consequence is often falsely drawn from these considerations. Such importance is attached to the intentional explanation of human behaviour and activity, and indeed to the variability of social context, that the large-scale explanation of educational practice is rejected entirely. A sharp contrast is drawn between the kind of evidence which pertains to the explanation of physical events (and included in that would be the successful intervention of drugs in the treatment of diseases) and the kind of evidence which pertains to the explanation of human behaviour. Favoured by the former, but not by the latter, would be the large-scale and carefully matched experimental and control groups, in which a particular intervention within the experimental group (all else being held equal) would demonstrate its causal significance. Certainly, such large scale experiments are seen to be the way forward by some in advancing our knowledge of educational improvement. And there are examples of such interventions in research into early learning (see Sylva and Hurrell, 1995, into the effectiveness of Reading Recovery
and the phonological training of children with reading problems). However, the critics point to the failure of such evidence to address the particularities of the social situations which are meant to be explained. And that failure is seen to be at base philosophical – the adoption of what are often referred to as positivism which has no place in our understanding of human beings and social institutions. There can be no ‘science of man’ – the title of a paper by the once most prominent logical positivist, A. J. Ayer.

This is surely a mistaken conclusion. It commits what I refer to as the ‘uniqueness fallacy’. It is correct to point to the uniqueness of each individual, since he or she is defined partly in terms of the particular way in which the world is seen and appreciated (no one can have exactly my thoughts and feelings). Similarly, it is correct to point to the uniqueness of each social group or society, reflected in the social rules and expectations which distinguish that group. But although each person or each society might be unique in some respect, it is not the case that each is unique in every respect. I am unique in that no one shares the same life history, but I am not unique as an Englishman, as a university professor, as a writer, as an allotment holder. And, in all these things, I can, within certain parameters, be predicted, under normal conditions, to behave in a certain way.

Furthermore, what enables me to understand (if only to some extent) people within very different societies from my own is that all human beings, whatever their differences, have certain propensities, desires, needs and wants in common. There is such a thing as a recognisable human form of life which enables us to make predictions, even though in particular cases the predictions may be wrong – the person consciously bucks the trend. But even the exceptions can be understood in the light of further explanation that helps us to make sense of the situation. And ‘explanations’, by their very nature, put the unique case into a wider framework in which the uniqueness diminishes somewhat. A person fails to act as predicted because he was ambitious for a specific acknowledgement, but ‘ambition’ is a recognisable human motive. To say that someone acted out of ambition is to place his actions within a wider explanatory framework.

Therefore, to contrast so starkly the large scale explanations of human behaviour, characterised by predictors of what will happen (having arrived at such a position through randomised control experiments), with the uniqueness of the individual human condition, which escapes any such pigeon-holing, is a false dualism. Much is predictable about human behaviour. And key interventions can be identified which, generally speaking, will lead to certain consequences. To draw different conclusions is to commit the uniqueness fallacy.
But of course one needs to be very careful in spelling out the conditions in which the intervention is likely to make a difference. Such conditions might refer to the particular kind of institution or social arrangement. An intervention in a highly selective system of schooling might have little effect in a non-selective system. The literacy hour might be effective in certain teaching environments and not in others. The Cochrane ideal was not to determine professional practice but to inform it. The teacher, aware of what generally speaking is likely to be the case, may well exercise professional judgement about the circumstances, which are judged to be sufficiently different from the norm as to create an exception to the general rule.

**Means and ends**

The concern for evidence-based policy and practice arises within a climate of ‘improvement’, ‘raising standards’, ‘making schools more effective’. Knowledge is required of ‘what works’. To do this, so the argument goes, there is a need to set targets, as specific as possible. These are the goals to be aimed at, the ends to be striven for. It seems plausible to argue that you cannot be very effective until you know exactly where you are going. Only then can you focus your energy and effort on reaching your goals. Having established those targets, the school or the local authority or the government can then discover (by the most appropriate empirical enquiry) the way in which those targets can be met. Such investigation relies upon unambiguous and clear targets. And it requires rigorous research into the most effective means of hitting those targets.

Within such a climate, there has been in the last decade a massive expansion of research into school effectiveness – the characteristics of a school and its leadership which will ensure ‘success’. Success is spelt out in terms of very precise targets (such as a given proportion of students attaining grades at GCSE and A Level). Similarly, effective teaching (clearly essential to the effective school) is defined in terms of pupil performance which can be precisely measured. With systematic gathering of evidence, one might develop a science of effective teaching (see, for example, Reynolds, 1998). Once the government or whoever is assured, on the basis of rigorously conducted experiments, of the right interventions to make, then it will put in place the right mechanisms for ensuring higher performance against the agreed standards. And, indeed, teachers will then receive payment which is performance related.

It is within this climate that a major authority on evidence-based education policies (Slavin, 2000) confidently writes about ‘transforming educational practice and research’ and refers with approval to the various government initiatives which have adopted
‘experimental-control comparisons on standards-based measures’. For example, the Bush administration’s ‘No child Left Behind’ mentions scientifically based research 110 times – ‘rigorous, systematic and objective procedures to obtain valid knowledge … using experimental or quasi-experimental designs, preferably with random assignments.

Within the now prevalent managerial discourse (a discourse of ‘performance indicators’ and ‘audits’, of ‘curriculum delivery’ and ‘efficiency gains’, of ‘targets’ and ‘value addedness’, of ‘clients’ and ‘stakeholders’), the means / end model of educational planning and engagement seems almost self-evidently correct. There is a logical separation of the ‘ends’ of education form the ‘means’ of achieving those ends. The connection is purely contingent, a matter solely of empirical investigation. And in the educational encounter, the teacher is the expert (hopefully on the basis of the right evidence) in knowing what ‘means’ will most effectively attain those ‘ends’. The teacher’s expertise lies not in the deliberations over the ‘ends’ themselves.

Such a language, which lends itself to a particular understanding of evidence-based policy and practice, is superficially plausible, but is a quite impoverished way of talking about and understanding education, for the ‘ends’ are more often than not embedded within the ‘means’. The way in which one analyses a poem is not assessed in terms of being the most effective way of attaining goals, logically distinct from the reading and the analysis of the poem. The goal, end or purpose shapes the way in which the teacher teaches – it is captured and ‘shown’ in the very act of teaching. Teaching is a transaction between the teacher and the learner, not the delivery of something to the learner. An educational practice embodies the aims and values; it is not something distinct from them. Indeed, to ask for the aims of such a transaction is to ask for the vales which the transaction embodies. There may well be ‘spin-offs’ from teaching MacBeth, but the main educational purpose lies in the engagement with a valuable text. The language of ‘engagement' with a text, of ‘transaction between teacher and learner’, of ‘intrinsic value’ of an activity, of ‘struggle to understand’, of ‘personal enrichment’ seems inimical to the language of targets and of standardised performance indicators or of generalised conclusions drawn from systematic interventions. And it is this view of an educational practice which underpinned the seminal work of Stenhouse (1975), referred to and developed by Elliott in this volume.

**Conclusion**

There are different levels at which one can examine and appraise evidence-based policy and practice in education. Educational policies aiming to improve the quality of learning
and to increase the number of people who successfully participate in education at different phases need evidence to show that one policy rather than another will make things better. Teachers, in the myriad judgements they make every day, would be more professional in those judgements if these were based upon the accumulated evidence from their own practice and from that of the profession as a whole. Of course, that is what they claim to do. Staff room talk is as much about what has worked, or about advising others in the light of what has been seen to work, as it is about anything else. And so at one level there cannot be much dispute about the idea of evidence-based policy and practice. Teachers, ministers and civil servants give reasons for what they do and those reasons necessarily call upon evidence.

The advocates of evidence-based policy and practice, however, argue that the gathering and the application of evidence has not been rigorous enough. It lacks the systematic investigation, indeed the scientific rigour, which has transformed other areas of public life. Educationists are chastised for their failure to search for evidence systematically enough. Thus Slavin (ibid. p.16) states

> At the dawn of the 21st century, education is finally being dragged, kicking and screaming, into the 20th century. The scientific revolution that utterly transformed medicine, agriculture, transportation, technology and other fields early in the 20th century almost completely bypassed the field of education. If Rip Van Winkle had been a physician, a farmer, or an engineer, he would be unemployable if he awoke today. …It is not that we have not learnt anything since Rip Van Winkle’s time. It is that applications of the findings of educational research remain haphazard, and that evidence is respected. only occasionally, and only if it happens to correspond to current educational or political fashions.

The problems arise, therefore, not over the need for evidence in the adoption of policies or in the improvement of practice, but, first, over what is to count as evidence, second, over the extent to which the scientific rigour in some areas are equally applicable to educational policy and practice, and, third, over whether there is something so distinctive and peculiar about an ‘educational practice’ that there are strict limits to the relevance of the means / end model of educational improvement and effectiveness.

Thus, as I argued, evidence is of different kinds relative to the form of discourse through which a problem is being addressed. For some (and there are hints of this in the contribution to this volume by Hodkinson and Smith), such an admission leads to the sort
of relativism which makes a nonsense of the evidence-based movement. But that does not follow. The different forms of discourse are not arbitrarily developed; they are the best window we have upon the world; and they have built into them the criteria of appropriate evidence without which one would not be able to engage in any intelligible argument – including arguments about evidence-based policy and practice.

Given the range of possible discourses about education, then the danger lies in the imperialism of any one form of discourse, together with its distinctive notion of evidence. Two false consequences are frequently drawn from this, exemplified in the contributions to this book. On the one hand, a narrow and thus too demanding a notion of evidence is adopted, thereby excluding, as irrelevant or as not rigorous or as arbitrary, deliberations about educational policy and practice. On the other hand, in recognising the distinctively practical, context bound and value-laden nature of educational deliberations, many will reject completely the large scale experimental search for evidence. Thus is created the false dualism between the quantitative and qualitative approaches to research, which has caused so much damage (see Pring, 2000, where this point is developed much more thoroughly).

There are three conclusions that need to be drawn from this as we look to the future.

The first is that evidence-based policy and practice need to look much more carefully at the different kinds of evidence which legitimately enter into educational deliberations at the policy and professional practice levels. Notions like deliberation, personal and craft knowledge, as well as the different kinds of evidence which enter into educational discourse should be examined critically. It is important to explore what ‘systematic’ means within these different kinds of appeal to evidence.

Second, despite the rather eclectic nature of educational discourse, there are lessons to be learnt from the insistence by the advocates of evidence-based policy and practice for the more rigorous search for evidence. These are the constant attempt to synthesise and reconcile the different research findings, the search for the logical connection between conclusions drawn from different kinds of research, the assessment of the degree of reliability of the research for future policy and practice, the evaluation of the conclusions in the light of the explicitly reported data and methodology, the reporting of the research in clear and focused way.
Third, the political and often highly charged context of educational research (reflected so powerfully in Gallagher’s paper) needs to be recognised. It cannot be wished away. And that political context invades not only the policies and practices themselves, but also the different philosophical advocacies of different sorts of research. However pure and systematic the research should ideally be, it never will be like that. Slavin acknowledges the way in which ‘educational and political fashions’ affect the research, preventing the scientific objectivity which he is so anxious to promote. But his own paper, in persuading us of his position, is not without its own political rhetoric to get the point across.

References


