PROMOTING EVIDENCE-BASED EDUCATION: THE ROLE OF PRACTITIONERS

Round table presented at the
British Educational Research Association Annual Conference,
Cardiff, September 2000

Robert Coe, Carol Fitz-Gibbon and Peter Tymms

Curriculum, Evaluation and Management Centre, Durham University
Mountjoy Research Centre 4, Stockton Road, Durham   DH1 3UZ

Tel: 0191 374 4504; Fax: 0191 374 1900; Email: r.j.coe@dur.ac.uk
http://www.cem.dur.ac.uk/ebeuk

Abstract

A number of recent initiatives from Durham University's Curriculum, Evaluation and Management Centre have sought to involve teachers in creating, accessing and applying evidence about what works in their practice. The 'gold-standard' of evidence in this context is taken to be multiple replications of small scale, randomised controlled trials of feasible interventions in real-life settings. The aims, form and progress of these initiatives will be reported, and a number of questions will be raised:

What do we mean by 'Evidence-Based Education'? How can it best be promoted? What kinds of research can teachers do? How good can it be? Can it genuinely contribute to knowledge? Is it a distraction or enhancement of teachers' core role? How do traditional models of Action Research fit with this approach? Is there an existing body of knowledge that can inform practice? How can teachers gain access to it? Under what conditions might such knowledge have an impact on practice?

Key words

Evidence-Based Education, RCTs, practitioner research

What do we mean by 'Evidence-Based Education'?

A short history of E-BE

The name 'Evidence-Based Education' is borrowed from Evidence-Based Medicine, defined as

> "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available

external clinical evidence from systematic research." (Sackett et al, 1996)

'Evidence-based' medicine traces its history back to nineteenth century Paris and beyond, but became fashionable only in the early 1990s. The formation of the Cochrane Collaboration in 1993 and, in the UK, the establishment of the Centre for Evidence-Based Medicine in Oxford in 1995 were important markers in its development. Parallels between the ways evidence was used in health care and in social policy areas like education began to be made at this time (eg by Hargreaves, 1996). The first in a biennial series of conferences on 'Evidence-Based Policies and Indicator Systems' was held in Durham in 1997 and again in 1999 (http://cem.dur.ac.uk/ebeuk; Constable and Coe, in press), with a third planned in July 2001. Debates about the meaning and application of 'evidence-based policies' began to be heard (Davies, Nutley and Smith, 1999; Davies, 1999). In July 1999, an initial meeting of the Campbell Collaboration, a younger sibling of Cochrane, was held in London. The Collaboration was formally established at a meeting in Philadelphia in February 2000, with the aim of "Preparing, maintaining and promoting the accessibility of systematic reviews of the effects of social and educational policies and practices" (http://campbell.gse.upenn.edu).

Our definition

By 'Evidence-Based Education' we mean the support for and promotion of practices and policies that are based on good evidence about their effects (ie costs and benefits). Of course, the question of what constitutes 'good evidence' is somewhat controversial. For us, there are two main elements to this. Firstly, it is important to note that actions like giving advice, advocating or requiring a particular practice or implementing a policy at some level are all interventions. And in order to know the effects of an intervention one has actually to have *intervened* in some way, not merely observed or described an existing situation. Secondly, for us to be confident that the benefits outweigh the costs, the intervention must have been well evaluated, ideally by a systematic review of multiple randomised, controlled trials (RCTs) of feasible interventions conducted in real-life settings. These two elements will be considered in more detail below.

Note first that we are not saying that all research in education should be limited to RCTs. There are many other kinds of research that may be just as 'relevant' to practitioners (Kennedy, 1999) or important in other ways (Hammersley, 1997). However, if research is to influence practices or policies, this can only be justified on the basis of sound knowledge about their likely effects, ie it must be 'evidence-based'. Alternatively, if educational researchers are happy to refrain from giving advice to practitioners or policy makers, then they may be free to conduct whatever kind of research they choose. Nevertheless, we would argue that RCTs have been somewhat out of fashion in many educational research circles over the last 20 or so years, and that their usefulness probably has been underrated.

We should also note that sometimes there is no good evidence available, but we may still have to act. In this case an 'evidence-based' approach would be to seek out and act on the best evidence that was available, but perhaps more importantly, to make it a priority to create the kind of evidence that might be considered a secure basis for action.

Intervention, not description

It does not necessarily follow that because successful schools tend to do 'X', that by encouraging 'X' one will make schools more successful. Put thus it seems entirely obvious, yet statements that effectively confuse description and intervention in this way are depressingly common within educational research and policy. It is arguable, for example, that the whole of school effectiveness research is founded on this error (Coe and Fitz-Gibbon, 1998). A recent British example is the Hay-McBer report on 'teacher effectiveness' (DfEE 2000), for which the government allegedly paid £4 million (Barnard, 2000). The report describes what effective teachers do and makes the (generally implicit) assumption that by seeking to adopt these characteristics, teachers will become more effective. Since teachers' pay will depend on their ability to demonstrate the characteristics it describes, they are sure to be widely adopted. But whether this will lead to any real improvement in teaching seems very much open to question.

Of course, it may well be true that by copying good schools, other schools will indeed become better. However, it may equally be that their characteristics cannot simply be adopted at will, or that even where they can, adopting them will not lead to the improvements sought. At best, this is a hypothesis that deserves, but cannot be judged without, proper testing. In no way does an argument of this kind constitute 'good evidence'. An evidence-based approach requires a higher standard of justification.

Evaluation, not common sense

There is something of a tradition in social policy formation that if something can be plausibly argued to be beneficial then that is justification enough, particularly if it is thought likely that voters will approve. Recent government attempts to become 'Evidence-Informed' may be seen as a welcome reversal of this tradition, requiring more than just common sense to back up their policies. The tradition, however, is widespread throughout education, from the statutory requirements of governments, to the advice given by educational experts and consultants, to the policies adopted by headteachers, heads of departments and classroom practitioners. When 'policies' at all these levels have to be justified, it is seldom done in terms of any kind of rigorous evaluation of their effects.

Despite this reliance on common sense, there are plenty of examples of plausible and well meant interventions whose effects have been far from what was expected or intended (eg McCord, 1978). Common sense is no substitute for research (Tymms, 1999). Moreover, poor research is no substitute for good research. In a review of evaluations of social policy interventions, Boruch (1997, p69) found that, "Declarations that a program is successful are about four times more likely in research based on poor or questionable evaluation designs as in that based on adequate ones." In order for practices and policies to be described as 'Evidence-Based', therefore, it really is necessary for them to have been evaluated properly.

What does 'well evaluated' mean?

There are a number of characteristics of an evaluation that may be expected to determine the quality of the evidence it generates. Perhaps the foremost of these is the kind of control used. If you do 'X' and find that some desired outcome improves, you need to be sure that it would not have improved just as much (or even more) had you not done 'X'. Ideally, you want to be able to quantify and compare the amount of improvement produced by 'X' with that produced by other interventions 'U', 'V' and

'W' (one of which might be simply to carry on as before). In this way you can assess the costs and benefits of each, and make an 'Evidence-Based' choice among them. This kind of comparison or control is generally provided by having two (or more) groups in the trial and treating them differently. Allocating people to groups *at random* is a simple way of guaranteeing that the groups are the same before you start the intervention.

In the context of medical interventions (from which the term 'Evidence-Based' is taken) there is strong evidence that evaluations using randomised designs give results that are different from those using non-randomised (or poorly randomised) allocation to treatments (Kunz and Oxman, 1998). However, in a meta-analysis of meta-analyses of the effects of psychological, educational and behavioural interventions, Lipsey and Wilson (1993) found that, overall, evaluations with randomised designs did not give appreciably different results from those with non-randomised designs. They did, though, find that effect size estimates from studies without a control or comparison group were substantially higher than those from studies with a control, and that published studies gave higher estimates than those that were unpublished. Interestingly, they found no systematic relationship between effect size and a study's rating of methodological quality.

The empirical evidence, therefore, suggests that evaluations without a control group probably do not constitute 'good evidence', but empirical support for requiring random allocation to treatments is rather more equivocal. Nevertheless, there are good theoretical reasons for preferring random allocation in terms of its reduction of the risk of a variety of threats to validity (Campbell and Stanley, 1963; Cook and Campbell, 1979). Above all, it is important that if there is a control group, it must be shown to be initially equivalent to the treatment group; random allocation is generally by far the most convincing way of doing this.

Another feature of evaluations that may influence our judgements about how valid they are is the range and nature of the contexts in which they have been carried out. A single experiment carried out in one school may provide some evidence of value to that school, but even then we would have more confidence in the results if we knew they had been repeated. The history of the natural sciences shows clearly that one-off results are not reliable until they are replicated.

If we want to generalise the results to other schools, we certainly need more experiments to be done. Social science phenomena are often so dependent on context, and school contexts are so varied, that generalising results is usually very difficult. Until we understand a phenomenon really well, we cannot be 100% confident that it will generalise to contexts in which it has not been specifically tested. Meanwhile, there will be room for debate about which contexts are sufficiently close: which will tell us more about what will happen in a school in inner-city Cardiff, a trial in inner-city Chicago or one in rural North Wales? Ideally, if an intervention is not universally effective, we should try to identify what features of a particular context determine whether or not it works. Where sufficient studies exist, good meta-analysis based on systematic review of results from multiple contexts can supply this kind of knowledge. Certainly, if we want to be able to describe a particular policy as evidence-based, then we must be reasonably sure that its effects can be generalised to all the areas it covers.

Isn't teaching rather more complex than that?

One of the arguments sometimes put forward against Evidence-Based Education is that activities like teaching are too complex, too dependent on the

particular context and moment, too much the result of subtle interactions among people – who are actively interpreting, responding to and shaping their environment, rather than simply reacting to it in a predictable fashion – for anything so crude and simplistic as a randomised controlled trial to be of any use. Certainly, teaching is complex, and this complexity may be so great that we will never usefully be able to predict behaviour, other than in very restricted circumstances. However, if that is the case, then it amounts to saying that there can never be an intervention whose effects are predictably beneficial. If that is so (and we would regard it as an empirical question whether or not it is), then the only evidence-based advice one could give would be to give no advice at all. Politicians and other self-appointed experts would have to leave decisions to be made at the local, situation-specific level. Teachers might welcome this lack of interference. On the other hand, if there are interventions that can be shown to be predictably beneficial in particular circumstances, then we think most teachers would want to know about them. Moreover, we believe that there are indeed examples of interventions that have had consistently positive results.

What have we done so far?

The main work of Durham University's Curriculum, Evaluation and Management (CEM) Centre is working with schools to provide comparative feedback on a range of performance indicators, notably on the 'value added' progress of their students. Over 7000 schools in the UK currently participate in one or more of our projects. The focus for these schools is on self-evaluation, based on sound evidence (see http://cem.dur.ac.uk). It may be, therefore, that these schools, who require (and pay for) good evidence about their own effectiveness, would be particularly receptive to the idea that policies at various levels should be based on equally good evidence. The CEM Centre's work in promoting Evidence-Based Education may thus be seen as a natural development from its work in performance monitoring. The need for evidence-based education to include both monitoring and experiments has been persuasively argued by Fitz-Gibbon (1996).

To this end, we have discussed and promoted the aims of E-BE with the schools in our projects (and others) and created an 'Evidence-Based Education Network' – a list of names of those who are interested in the ideas. Information about the Network can be found in our web site at http://www.cem.dur.ac.uk/ebeuk . Currently it contains about 400 names, by far the majority of whom are practising teachers. There have been two Newsletters (so far!) with information about developments and ideas.

The web site lists four main aims for the Network:

- To create evidence
- To disseminate evidence
- To promote a culture of evidence
- To campaign against 'non-evidence-based' policy and practice

The first two of these are concerned with involving practitioners in research, the second two relate to a more general promotion of the ideas of Evidence-Based Education. It is hoped that the Network will become a means of encouraging and supporting teachers' participation in and engagement with research. With these aims in mind, we have been able to hold three meetings in the last year for teachers to exchange ideas and see examples of the kinds of research we are keen to promote.

The first of these was the last seminar in a series of four on 'Evidence-Based Practices and Policies', held at the Royal Society in London on 18 November 1999. The seminar was publicised with the questions shown in Figure 1. It was attended by about 50 teachers and their feedback (which can be found on the web site) was extremely positive.

How can you *reliably* raise achievement?
What kind of research can tell us what works?
How do you get access to educational research?
How can teachers contribute to research?

Figure 1

The format of the day consisted of eight brief presentations by practising teachers about the research they had each done. The full programme can once again be found on the web site, but Figure 2 gives a summary of the research question addressed by each. Each of these teachers had conducted a small-scale investigation, usually a randomised controlled trial.

Finding out what works:

< Teaching French vocabulary: give students a specific strategy or let them choose?

< Anti-racist education: does it reduce racist attitudes?

< Can you teach general movement coordination?

< Do children learn best in the morning or afternoon?

< Can year 10 students benefit from teaching fractions to year 7?

< Year 9 Business Studies: do they learn best individually or in groups?

< Whole class teaching vs. individual work: which works better teaching French in a special school?

< Is group reading in the Literacy Hour better than individual help?

Figure 2

The success of this seminar prompted us to want to repeat it, and we were able to secure some funding from the Teacher Training Agency to run two further conferences in March 2000, one in Leeds and one in London. The format of these conferences also centred on presentations by teachers of their research, followed by some input from CEM Centre personnel on some more technical research issues. Once again, the conferences were a great success; they were attended by about 140

teachers and the feedback was very positive. Details can once again be found on the web site.

Involving teachers in research

There are a number of reasons for wanting teachers to take part in research. Detailed justifications for eight of these reasons are given below.

1. We need more research to be done

Some have argued that the main focus for evidence-based policy should be the systematic review of already existing research, rather than the creation of new studies (Chalmers 1999). Substantial amounts of public funding have recently been allocated to the collection, systematic review and dissemination of research evidence about effectiveness (eg £1.9m from DfEE for EPPI at Institute of Education, London; ESRC funding for EB Soc Sci Centre ????). Clearly, it is important to know the results of any relevant existing research if one wishes to judge the effects of a particular intervention. And if, after considering all the evidence, it seemed that sufficient knowledge was already available to support evidence-based action, then it would be a waste of time and resources to conduct further studies.

However, we would argue that there are very few – if any – areas of education in which such a level of knowledge exists. Certainly, there are plenty of questions of practice or policy where research seems to have almost nothing relevant to say. If more research is needed, then it can only be done if teachers are involved.

2. Only those who do the job can ask the right questions in the right ways

In any experiment, the question of what outcomes to record is, at least in part, a value judgement. For example, most school effectiveness research has tended to define 'effectiveness' in terms of student performance in rather narrowly conceived tests (eg basic mathematical skills or reading). Most teachers, on the other hand, would probably want a much broader range of outcomes to be included in any definition of effectiveness. Unless we record the effects of an intervention in terms of outcomes that are agreed to be important, our research will be useless. Reaching this agreement must follow a process of negotiation in which practitioners are involved.

Another example of a failure to record important outcomes is given by Oakley (1992) from the health care field, who questions whether the advice given to pregnant women not to smoke is necessarily beneficial. This advice causes many mothers either to feel guilty if they do smoke, or to lose an important coping mechanism and feel stressed if they try to give up. Given the existing stresses of pregnancy, particularly for those in relatively disadvantaged circumstances, she argues that it is not clear that the benefits for the baby outweigh the costs to the mother. Unfortunately, because almost all available studies have focused on strictly 'medical' outcomes, rather than the experiences of those involved, there is very little evidence with which to resolve this.

The relevance of the initial research question is also at issue here. Educational research has been criticised for being irrelevant to the needs of practitioners (eg Hargreaves, 1996). If this criticism is fair, then one way to address it might be to involve practitioners in the choices about what questions are researched.

3. Only those close to the outcomes can provide a rich and detailed understanding of them

There are a number of reasons why practitioners should be involved in the process of recording and interpreting the outcomes of an experiment.

Firstly, it is often the case that some of the important effects were not anticipated at the design stage. It was not intended to measure them, and unless you were there you might easily not have noticed a particular difference between groups.

Secondly, and contrary to popular misconceptions about experimental research, it is not just quantifiable or numerical outcomes that are important. Qualitative data can provide evidence of individual experiences and interpretation that are vital to understanding and evaluating the impact of an intervention. Self-evidently, the experiences and interpretations of those involved in the experiment can only come from those involved. Detailed information about the context in which an experiment was conducted can also be crucial.

Third and finally, it may be that although many aspects of the results of an experiment can be presented objectively, their ultimate interpretation must remain subjective. Certainly, this is the approach taken in all the CEM Centre's monitoring projects. We feed back the data, together with the necessary support, and leave the schools to interpret it. Most experienced users of monitoring systems would say that the data do not speak for themselves, that they cannot be interpreted without a detailed knowledge of the context in which they were produced; data provide questions, not answers. Perhaps the evaluation of interventions will prove to be the same.

4. Many teachers are already experimenting

Many teachers and schools already develop their practice continuously by trying new things. This kind of experimentation is at the heart of a 'problem-solving' approach to education. Moreover, if adequate feedback is available it can provide a very quick way to learn what works. Involving teachers in 'research', therefore, is just an extension and formalisation of existing good practice.

However, there is a bit more to it than this. It has already been argued that it is easy to get it wrong when trying to evaluate whether something has worked or not. It is hoped that the use of more rigorous forms of evaluation would reduce this danger. Also there is a good deal to be gained by the pooling of results from different teachers in different schools all of whom have experimented in much the same way.

5. Assimilation, not dissemination: other people's ideas don't have the same impact

Some have blamed the lack of impact of research on practice on the failure of researchers to communicate their findings well (Hargreaves, 1996). But practice is hard to change, teachers are busy. How much effect can dissemination alone have? For change to occur, new ideas have to be assimilated into one's practice, internalised and made automatic. It may be that a more active involvement in research, and engagement with its debates and ideas would be more likely to bring about this change than simple dissemination.

6. We must evaluate actual implementation, not just ideal policy

A vital but often neglected part of interpreting the results of an experiment is knowing what was actually done: ie the extent and quality of the implementation. If, as is common, the intended intervention was tweaked in some way, if only part of it was judged appropriate and the rest not used, or if it was adopted in a half-hearted

fashion, then one would need to know this in trying to understand what had caused any difference (or lack of it) in outcomes. In practice, all policies and programmes are modified or 'corrupted' by those who implement them. This is inevitable and proper – they must be made to work. If our aim is to improve real practice in some way, then it is important to evaluate something that is real, feasible and sustainable, not just an idealised policy. This kind of information about how the interventions were implemented is often best provided by an insider.

7. Only multi-site trials can give generalisable findings

We have already argued that evidence must be drawn from a range of contexts. Involving practitioners from a range of different schools, neighbourhoods or regions in a single trial provides a way of achieving this.

8. The process of doing the research is itself valuable

It is clear that the teachers who have taken part in these research projects have benefited from doing so in a variety of ways, and we believe that many other teachers would also find this valuable. In accordance with our belief in evidence-based practice, we have also tried to evaluate the conferences we held and the feedback from the teachers who took part was very positive. However, at the moment this belief must retain the status of a conjecture, since we do not know how the benefits of taking part in this kind of research would compare with other forms of training or support.

Discussion

How can it best be promoted?

We have described what we have done. Are there other things we could be doing?

What kinds of research can teachers do? How good can it be? Can it genuinely contribute to knowledge?

Quality of research funded by TTA and others??

Is it a distraction or enhancement of teachers' core role?

Teachers should be allowed to get on with teaching?

How do traditional models of Action Research fit with this approach?

Is there an existing body of knowledge that can inform practice? How can teachers gain access to it?

Under what conditions might such knowledge have an impact on practice?

References

Barnard, N. (2000) 'Revealed: the ideal teacher'. *Times Educational Supplement*, 16 June, 2000, p5.

Boruch, R. F. (1997) Randomized Experiments for Planning and Evaluation: a practical guide. London: Sage.

Campbell, D.T. and Stanley, J.C. (1963) *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin Company.

Chalmers, I., (1999) 'Infrastructure for international collaboration to prepare and maintain systematic reviews of educational interventions'. Presentation at a seminar on Evidence-Based Practices and Policies, at the Royal Society, London, 19 May 1999 (http://cem.dur.ac.uk/ebeuk/EBPP2.htm).

Cook, T.D. and Campbell, D.T. (1979) *Quasi-Experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.

Davies, H.T.O., Nutley, S.M., and Smith, P.C. (1999) 'Editorial: What Works? The role of evidence in public sector policy and practice' *Public Money and Management*, 19 (1) 3-5.

Davies, P. (1999) 'What is Evidence-Based Education?' *British Journal of Educational Studies*, 47 (2), 108-121.

DfEE (Department for Education and Employment) (2000) 'Research into Teacher Effectiveness: A model of teacher effectiveness'. Report by Hay McBer to the Department for Education and Employment, June 2000.

Fitz-Gibbon, C.T. (1996) *Monitoring Education: indicators, quality and effectiveness*. London: Cassell.

Hammersley, M. (1997) 'Educational Research and a Response to David Hargreaves' *British Educational Research Journal*, 23 (2) 141-161.

Hargreaves, D.H. (1996) 'Teaching as a Research-Based Profession: Possibilities and Prospects'. Teacher Training Agency Annual Lecture 1996.

Hargreaves, D.H. (1997) 'In defence of research for evidence-based teaching: a rejoinder to Martyn Hammersley' *British Educational Research Journal*, 23 (4) 405-419.

Kunz, R.H. and Oxman, A.D. (1998) 'The unpredictability paradox: a review of empirical comparisons'. *British Medical Journal*, 317, 1185-90.

Lipsey, M.W. and Wilson, D.B. (1993) 'The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation from meta-analysis'. *American Psychologist*, 48, 12, 1181-1209.

McCord, J. (1978) 'A thirty year follow-up of treatment effects'. American Psychologist, 33, 284-9.

Oakley, A. (1992) *Social Support and Motherhood*. Oxford: Basil Blackwell.

Sackett, D.L., Rosenberg, W., Gray, J.A.M., Haynes, R.B. and Richardson, W. (1996) 'Evidence-Based Medicine: What it is and what it isn't' *British Medical Journal*, 312, 71-72.

Tymms, P. (1999) 'Homework, Common Sense, Politics and the Defence of Research: Peter Tymms to BERA Membership' *Research Intelligence*, December 1999 (70) 22-23.