



## *Systematic Reviews of Educational Research: does the medical model fit?*

**JENNIFER EVANS**, *Institute of Education, University of London*

**PAULINE BENEFIELD**, *National Foundation for Educational Research, Slough*

**ABSTRACT** *There has been a recent increase in interest in the research review as a method of presenting cumulative data about the effects of educational policies and practices. This is part of a wider movement in 'evidence-informed policy-making' espoused by the current Government. In part, the interest has been sparked by the perceived success of the Cochrane Collaboration in medicine, which has set up a framework for conducting and verifying systematic reviews and meta-analyses of random controlled trials of medical interventions. A pilot project to apply the methods of systematic review has been carried out at the National Foundation for Educational Research. The subject of the review was 'Strategies to Support Pupils with Emotional and Behavioural Difficulties in Mainstream Primary Classrooms'. The article describes the process of the review, the adaptations of the 'medical model' to educational settings and discusses some of the implications of these for researchers and policy-makers.*

### **Introduction**

The current UK Government has made research a key building block of its approach to policy formulation and evaluation across the range of public services. *Evidence-based* or *evidence-informed* policy and practice are the new buzzwords in this area. David Blunkett, the Secretary of State for Education and Employment (DfEE) stated, in a press release:

Social science should be at the heart of policy-making. We need a revolution in relations between government and the social research community—we need social scientists to help to determine *what works and why*, and what types of policy initiatives are likely to be most effective. (DfEE, 2000, my emphasis)

He went on to report that the DfEE would be spending over £10.4 million on research in 2001–02, almost double what was spent in 1997. He called for three types of research; that which:

Received 16 November 2000; resubmitted 30 April 2001; accepted 4 June 2001.

- gives a coherent picture of how society works; the main forces at work, and those which can be influenced by government, such as inter-generational poverty, low aspirations, employability, participation in society or exclusion;
- evaluates policy initiatives and systematically reviews existing evidence;
- and, if we are to encourage a more open debate of ideas, we also need 'blue skies' research which thinks the unthinkable. If academics do not address this, it is difficult to think of anyone else who will. (DfEE, 2000)

However, echoing the words of Hargreaves's (1996) critique of educational research, he also stated that:

There is a widespread perception, both within and beyond government, that too much social science research is inward-looking, piecemeal rather than helping to build knowledge in a cumulative way, and that issues for research are too 'supplier-driven' rather than focusing on the key issues of concern to policy-makers, practitioners and the public at large, especially parents. (Blunkett, 2000)

This ambivalent attitude to social research, as on the one hand vital for policy-making, and on the other self-serving and inward-looking, has permeated debates on research during the late 1990s and into 2000. The critique provided by Hargreaves stimulated a huge debate about the methods and purposes of educational research (Hammersley, 1997; Foster, 1997; Gipps, 1997; McIntyre, 1997; Foster & Hammersley, 1998; Coe & Fitzgibbon, 1998; Tooley & Darby, 1998; Ruddock & McIntyre, 1998).

The Government's response to this has been to set up a number of structures to target research into those areas in which it has a current interest and to support the research community to deliver the models of research which it judges to be necessary to achieve this endeavour. The focus of much of the activity in this area has been on 'what works' in terms of social policy, i.e. on the *effectiveness* of social interventions. As Ann Oakley has pointed out (Oakley, 2000), evidence of effectiveness is often best provided through some form of experimentation and this has been the dominant mode of research within medicine, but has been resisted to a large extent within social science. Oakley asks:

Why did experimental methods appear only to belong to natural science and to doctors? Why did they not seem to be owned equally by social science and other groups of professionals—for example, social workers, teachers and those involved in criminal justice and crime prevention? What was so special about those professions that made them immune to the need to show that their interventions in other people's lives worked and did more good than harm? (Oakley, 2000, p. 19)

While this is overstating the case for the lack of experimental research in social science (especially in the USA, where it has been more prevalent [see Oakley, 1998a]), in the UK, there has been an evolving tradition in evaluation research, of using illustrative case studies and the thick description of qualitative interviews, rather than experimental designs, to evaluate interventions. Thus, the questions addressed were not so much 'What works?' but 'How does it feel?' The latter is an important question, but does not necessarily tell us anything about whether the aims of the intervention (reductions in offending, gains in learning, etc.) have been achieved.

Another criticism of educational research made by Hargreaves (1996), and taken up by David Blunkett in his speech to the Economic and Social Research Council (ESRC) (Blunkett, 2000), was that it was not cumulative in nature, and therefore did not build

a body of knowledge which could inform policy-makers and practitioners. Again, in contrast to the health sector, where systematic reviews of research have been developed through innovations such as the Cochrane Collaboration (Sheldon & Chalmers, 1994), there was no accepted framework for presenting the cumulative results of educational and other social science research. Again, in the USA, there have been a number of efforts in this direction, such as Slavin's work on research synthesis (Slavin 1986, 1995) and Glass's work on meta-analysis (Glass, 1976; Glass *et al.*, 1981). More recently, standard textbooks on research synthesis have been produced—for example, Cooper (1998) and Cooper & Hedges (1994). Most of this work is based on the use of quantitative data and statistical analysis techniques, although there has been some work done on synthesising findings from qualitative studies (Noblitt & Hare, 1988).

### What is Meant by 'Systematic Review'?

MacDonald (2000) suggests that systematic reviews of research:

entail a series of techniques for minimising bias and error, primarily through the use of *protocols* which state, prior to the review being undertaken, what the criteria are which will guide the review, search strategies, inclusion and exclusion criteria, standards of methodological adequacy, the precise definition of the intervention in question, unbiased estimation of aggregate effect, and so on. (MacDonald, 2000, p. 131)

The key features of systematic reviews, then, are: an explicit research question to be addressed; transparency of methods used for searching for studies; exhaustive searches which look for unpublished as well as published studies; clear criteria for assessing the quality of studies (both qualitative and quantitative); clear criteria for including or excluding studies based on the scope of the review and quality assessment; joint reviewing to reduce bias; a clear statement of the findings of the review.

The key difference between systematic reviews and the more common narrative or academic reviews is that the former are explicit about all the above (Slavin, 1995). Narrative or academic reviews tend to be less focused and more wide-ranging in their scope (see, for example, Hart & Nolan, 1999), less clear in reporting their search strategies or their criteria for including or excluding studies (for example, Gillborn & Gipps, 1996), have less explicit criteria for assessing the quality of the studies included (for example, Hallam & Cowan, 1998) and do not generally make explicit their methodology for reviewing the studies included.

Bassey (2000) has distinguished between 'academic' and 'user' reviews on the basis of the audience for their findings. However, it is not so much the audience, but the purpose of the review which is the key feature. That is, what question is the review asking? If it is, as in Hart & Nolan (1999), 'What research was carried out on environmental education between 1993 and 1999?', then the broad scope and general conclusions are to be expected and the audience is likely to be academics and researchers. However, if the question is more focused, for example, 'Is homework important for increasing educational attainment?' (Hallam & Cowan, 1998) or 'Can school improvement overcome the effects of disadvantage?' (Mortimore & Whitty, 1997), then one would expect a more explicit statement of the types of evidence used to inform the review and of the criteria used to evaluate the evidence upon which the final conclusions are based. The audience for such a review is likely to be policy-makers and practitioners (and maybe parents), as well as researchers and academics. However,

the style of many of these more focused reviews is narrative and does not give the reader sufficient information upon which to form a view about the reliability of the conclusions. The key distinction, then, between narrative (or 'academic') and systematic (or 'user') reviews lies in the ways in which evidence is selected and evaluated and the explicitness with which this is reported (Slavin, 1995).

### **Evidence-based Policy and Practice**

This new terminology, derived from what has been happening in health over the last 10–15 years, fits neatly with the current focus on 'outcomes' and 'value for money' which underpins much government policy. This has been influenced by the market reforms in health and education delivery during the 1980s and 1990s, when there has been a much sharper focus on 'effectiveness' as a criterion for judging the performance of schools and hospitals and those who work in them. This current concern is exemplified by two recent research tenders from the DfEE, one of which required researchers to evaluate whether spending money on school buildings improved pupil performance and another whether increasing the administrative support to teachers (i.e. more help with photocopying, collecting dinner money etc.) would result in increases in pupil attainment.

Nevertheless, even though there may be some policy decisions where it would be very difficult to find evidence of effectiveness in terms of increases in pupils' test scores, many crucial policy decisions, in all areas of the public sector, require good evidence of their potential effectiveness. One way of assembling this evidence is the use of systematic reviews of research. A recent publication edited by Davies, Nutley & Smith (2000) makes out a clear case for this approach. Chapters cover health care, education, criminal justice, social care, welfare policy, housing, transport and urban policy. In each case, it can be seen that policy development is often incremental and based on custom and practice, rather than on a rigorous review of the evidence for effectiveness. However, in their review of the policy-making process, Nutley & Webb (2000) acknowledge that, despite David Blunkett's enthusiasm for social science research, policy-making is rarely a rational process, and evidence of what works is only one of the considerations that policy-makers have to take into account when making decisions. Nutley & Webb quote Weiss (1979), who described the various ways in which policy-makers might use research evidence:

- Knowledge-driven model—derives from natural sciences. The fact that knowledge exists sets up pressures for its development and use.
- Problem-solving model—involves the direct application of the results of a specific study to a pending decision.
- Interactive model—researchers are just one set of participants among many. The use of research is only one part of a complicated process that also uses experience, political insight, pressure, social technologies and judgement.
- Political model—research as political ammunition, using research to support a pre-determined position.
- Tactical model—research as a delaying tactic in order to avoid responsibility for unpopular policy outcomes.
- Enlightenment model—the indirect influence of research rather than the direct impact of particular findings in the policy process. Thus the concepts and theoretical perspectives that social science engenders pervade the policy-making process.

This broader view of the uses made of research might lead researchers to feel that their attempts to provide evidence about social phenomena or about the effects of policies and practices may be undermined by the realities of policy-making. However, as the example of health care has shown, through the Cochrane Collaboration, it is possible for researchers to assemble and review a body of evidence on a range of topics, which can be a guide for practice and a stimulus for policy development.

### **Systematic Review and Critical Appraisal in Health Care**

Over the last decade or so, there has been a major revolution in the field of health care which has been particularly influential in Canada and the UK: that is the movement towards evidence-based health care. Most health practitioners would support the principle that their practice should be based on sound scientific evidence. However, the problem had been that practitioners based their practice on idiosyncratic interpretations of idiosyncratic selections of the available evidence (rather than on an objective interpretation of *all* the available evidence), and important new evidence was not being incorporated into practice in an efficient and timely way. Inevitably, this led to a large diversity of practice, a significant proportion of which could be shown to be bad or suboptimal. Around the time the movement was started in the UK, it was estimated that only about 15% of the interventions in use in the National Health Service were supported by unequivocal scientific evidence (Department of Health, 1991). Meanwhile, practices which were well supported by evidence were often not being used. Together, these factors were having a profound effect on many tens of thousands of patients (Antman *et al.*, 1992).

The primary aim of the movement, then, was to encourage everyone to base their practice on the best available evidence of what was likely to work, and practitioners and policy-makers alike were encouraged actively to question their current policy and practice (why do we do this; is this the best way?), to review the evidence on which their rationales were based, and to determine if there were different practices which were better supported by the available evidence. In some cases, the best available evidence to address the issue of concern came from double-blind placebo random controlled trials (RCTs). In other cases, a different form of evidence was the most relevant and valid, or the only available, for addressing the question concerned. However, in either case, the imperative was to assemble and carefully review all of the most relevant evidence before coming to a judgement (Sackett *et al.*, 1997).

Probably the most important change brought about by the movement was that practitioners were encouraged to locate and appraise the evidence to answer their own questions themselves (North Thames Regional Office of the NHS Executive, 1998). No longer did their professional development depend on skimming a handful of journals and attending a few conferences to receive pre-digested and pre-organised information addressing issues which were not usually on their own main agenda. Instead, they could ask and answer their own pressing questions without waiting for someone else to ask the same question and provide the answer for them. In this way, research evidence was at last finding its way into practice in a timely way. Furthermore, at a time when budgets for professional updating and development were being progressively squeezed, to give people the knowledge and skills to be able to develop and update themselves in a competent way was seen as a very valuable contribution of the movement. At the unit level, some practitioners formed themselves into practice review groups which reviewed current practice and formulated best-evidence-based protocols for their units. Even in the

most prestigious teaching hospitals, these groups found many current practices which did not stand up to close scrutiny.

Although this activity was very relevant and brought enormous benefits, it could be very time-consuming and, despite the effort which has gone into pre- and post-registration training, most practitioners still lacked the knowledge and skills needed for carrying out competent systematic reviews of the literature. For these reasons, two other very important developments were the establishment of the Cochrane Centre in Oxford (part of the International Cochrane Collaboration) and the NHS Centre for Reviews and Dissemination in York, both of which facilitated and encouraged the carrying out of 'gold standard' literature reviews in key areas of practice, the results of which could be used to inform practice in a wide range of settings. The findings of these reviews are easily accessible online.

### **Will it Work in Education?**

It could be argued that the culture, training methods and organisational structure in the education service would make it difficult for a model derived from medicine to be transplanted and not to be rejected (see Pring, 2000). Many policy and practice initiatives in education are 'top-down' and mandated by Government, to an extent that they are not in the health service. Doctors have traditionally been seen as autonomous professionals who make clinical judgements about the most appropriate forms of treatment. Furthermore, the initial training and continuing professional development (CPD) of doctors and other professionals in the health service places a much greater emphasis on research than does the training and CPD of teachers.

However, as Davies (1999) has argued, it is easy to overstate the differences between the two services. Central government control, whilst not mandating specific treatments or interventions in health care, is nonetheless requiring similar sorts of accountability for outcomes and providing similar frameworks for reporting effectiveness to those demanded from education. There is a continuing debate about 'teaching as a research-based profession', following Hargreaves's address to the Teacher Training Agency (TTA) (Hargreaves, 1996), which may lead to changes in the ways in which teachers are trained and their subsequent involvement in and use of research. A survey for the TTA (Everton *et al.*, 2000) indicated that there is a need for the concerns of teachers to be given greater weight when research agendas are set and funding allocated, but also that, if research is to influence classroom practice, teachers should be given more opportunities to develop their skills in conducting and appraising research.

Recent developments in supporting the evidence-based approach in education include the setting up of the DfEE Centre for Evidence-informed Policy and Practice in Education (EPPI-Centre). The Centre has previously been conducting reviews in health, health education and social interventions. The EPPI-Centre's aims are: to provide methods and tools for carrying out systematic reviews of educational research; to set up an online bibliographic register of research; to provide training and practical support for groups of teachers and others in education to carry out research reviews; and to disseminate the findings of reviews.

Similar developments are taking place internationally with the setting up of the Campbell Collaboration, which is mainly based in the USA, but has international (including UK) participants. It is also involved with social work and criminology interventions. Campbell takes its inspiration from the Cochrane Collaboration in Health and will provide quality criteria and protocols for the carrying out of reviews of

educational research. In the UK, the ESRC has funded a Coordinating Centre for Evidence-based Policy and Practice at Queen Mary and Westfield College, headed by Professor Ken Young, the remit of which is to support interdisciplinary reviews of research across public policy domains.

All this activity led the National Foundation for Educational Research (NFER) to set up a working party to take forward the work of systematic review in the Foundation and to sponsor a pilot systematic review of research in one area of educational intervention. In the course of carrying out the review, we approached the EPI-centre (as it was then) to find out more about their methodology for reviewing studies in health education and health promotion. As the new DfEE centre (the EPPI-Centre) was just being set up, they agreed to work with us on the review to see how a model developed for reviews of health promotion would fit with the current available research in education and what adjustments might have to be made in order to take this type of review forward in education. The principles which we attempted to work to were those outlined earlier for systematic reviews, viz:

- a clear specification of the research question to be addressed;
- systematic and exhaustive searching for studies;
- clear criteria for including and excluding studies;
- assessments of methodological quality of studies;
- strategies to reduce bias in selection and reviewing; and
- transparency of the methodology for carrying out the review.

### **Review of Interventions to Support Primary-aged Pupils with Emotional and Behavioural Difficulties in Mainstream Primary Classrooms (the EBD Review)**

The long title of this review reflects an attempt to focus the research question to be addressed. The area of Emotional and Behavioural Difficulties (EBD) is broad, and interventions or adaptations can be at the level of the individual, classroom, school, home or in some specialised institution. EBD can also be associated with other learning difficulties or other more serious mental health problems. By confining the research to be reviewed to a specific age group (which was done to avoid the potentially confounding variable of adolescence) and to a specific setting (mainstream classroom), we attempted to delineate as clearly as possible the population and setting to be investigated. There was also potential overlap in the types of behaviours which interventions might be addressing. For example, we were not looking at studies which looked at interventions aimed at general discipline problems (i.e. for a whole class), but only those which were targeted specifically at pupils with EBD, although they may have effects on the whole class. However, we did include studies which were class-wide, but within that had specific components or specific targets for pupils with EBD. We excluded studies of interventions for children specifically diagnosed as having Attention Deficit Hyperactivity Disorder (ADHD), as many of these involved drug therapy, and thus were outside the remit of the class teacher.

The purpose of the review was stated as: ‘To describe and explain current knowledge about effective classroom practice related to EBD in order to guide professional practice’. The audience for the review was assumed to be primary classroom teachers, special educational needs coordinators (SENCOs), heads and deputies, advisers and educational psychologists. The criteria for the studies to be included were:

- focus on EBD;
- focus on classroom strategies;
- focus on primary age group;
- focus on children in mainstream; and
- focus on reports of research studies, or reports which have a research component.

We set up an advisory group of experts in the field of EBD or in systematic reviewing to help us clarify the field of study and to identify key researchers and authors in this field. The Centre for Reviews and Dissemination (CRD) has produced guidelines for the conduct of reviews, and we adopted these as a guide to the process (NHS Centre for Reviews and Dissemination, 1996).

### **Identifying Studies for Inclusion in the Review**

As we argued earlier, literature searching is a vital component of any systematic review. Details of the searching strategy are crucial aspects of the methodology section of the report, indicating the rigour of the search and thus the credibility that can be placed in the review's conclusions. It is these details which will be closely examined in the future if researchers or practitioners are attempting to understand why different reviews on the same topic area have reached similar or conflicting conclusions. This means that a systematic review of the literature must be conducted to an explicit and reproducible methodology. It is essential that reviews report their search strategies in terms of both the variety of sources used and the subject terms used for searching.

An initial stage of a systematic review would be to establish whether there have been any previous systematic reviews in the area. The medical and health care fields have an extensive evidence base, which is accessible through such national initiatives as the Cochrane Collaboration and its database of reviews known as the Cochrane Library. Educational research lacks such a centralised database, although one of the functions of the EPPI-Centre is to develop a database of research reviews.

The Cochrane Collaboration had begun work on assembling a database known as SPECTR (The Social, Psychological, Educational, and Criminological Controlled Trials Register). Responsibility for this has now been handed over to the Campbell Collaboration based at the University of Pennsylvania. SPECTR currently consists of approximately 10,000 randomised and quasi-randomised trials, but it is still at an early stage in its development.

The major sources used for searching for studies for the EBD review consisted of several electronic databases, in order to ensure comprehensiveness. It is an unfortunate fact that the databases that serve education are far less developed than those found in medical and health care research. MEDLINE is the major medical database and it has highly sophisticated searching strategies, allowing the insertion of quality strings (or evidence-based quality filters—EBQFs) into the searching strategy. These are designed to limit searches in various ways, either by type of intervention or type of publication.

In the education field, the most comprehensive database is the American Educational Resources Information Centre (ERIC), complemented by the newer International ERIC, covering the UK. At the time of carrying out our searches, this database still provided coverage of Australian and Canadian education, although it no longer does so. PsycLIT—the American Psychological Society's international database of the literature in psychology and related disciplines—also provides some education coverage. Other databases, Internet sites and personal contacts were used and will be detailed in the final report.

There were several predetermined search parameters which were used to decide the studies that were eligible for inclusion in the review: the age range had already been identified as 5–11 years; the timescale was specified as the past 20 years; the geographical limits were UK, USA, Australasia, Europe and Canada; and the types of literature were research-based.

The terminology for the literature search consisted of an 'Emotional and Behavioural Difficulties (EBD)' set of terms and a 'Classroom Strategies' set. Each set was then analysed and a much broader list of synonyms and related terms compiled, in order to broaden document retrieval. It is crucial in any systematic review that the search for primary studies should be extensive. (Primary studies are the independent sources of empirical evidence on which systematic reviews can be based.)

The same search strategies were adhered to as far as possible for all the databases using the predetermined search terms from the thesaurus. Searching with terms from the thesaurus helps to eliminate irrelevant records that may mention the word but do not focus on the concept. However, as this was a systematic review, it was essential that the search strategy located all the available relevant research, i.e. had high sensitivity. The thesaurus terms were, therefore, combined with free text searching—words which appear in the abstract, title and/or any other data field.

The searches were not limited by using the study design search terms, or filters, which are a recommended procedure in evidence-based medicine, as it was felt this would be too restrictive in the location of important study types.

The searching for additional studies began with the scanning of reference lists of relevant articles found through database searches. This is an invaluable strategy because it can help to identify further studies for consideration. It can also suggest new keywords and authors, which can lead to further searches for primary sources.

Where possible, key journals in the field were hand searched to identify articles that had been missed in the database searches. We supplemented our database searches with hand searches of 27 key journals. This is an extremely time-consuming exercise, but it did identify a further eight studies, some of which were more research-based than those found in the original searches. The importance of this process has been clearly shown by the Cochrane Collaboration, which identified approximately 30,000 trials that had not been appropriately tagged by MEDLINE in its first 5 years of operation.

We also explored why these studies had not been identified in the original searches by cross-checking the databases. All the studies were, in fact, present on the databases, but they had not been indexed using any of the classroom strategies set of keywords. They all, however, used at least one of the EBD set of terms.

Other types of literature that will need to be searched include grey or fugitive literature, i.e. material usually available only through specialised channels that does not appear on conventional databases. Grey literature can consist of documents such as final reports required by funding agencies, technical reports, discussion papers, booklets and opinion pieces from political parties. These are often only traceable with a great deal of detective work. Grey literature can also consist of totally unpublished studies. Comprehensive identification of such literature is hard to achieve, but much of it is now available on the Internet and it can also be indexed on databases such as SIGLE (System for Information on Grey Literature). This category may possibly also include undergraduate and postgraduate theses and conference proceedings. This type of material is indexed on the ERIC database and also on several more specialised databases, such as the Conference Papers Index, and Education-line, which is hosted by the British Education Index.

Although these databases constitute the major sources for searching, they have limitations:

- there can be a long time lag between when a study is completed and when it appears on the system;
- each database contains restrictions on what is allowed to enter the system, based both on the topic, and also on the type of literature. All the databases we used cover journal articles, but very few cover books or book chapters;
- there is no guarantee of the comprehensiveness within the topic—one estimate states that 40% of material that should be listed on MEDLINE can, in reality, be accessed only by looking through all the journals again, by hand. Indeed, some important and relevant journals may not be covered by the databases at all;
- entry of articles onto databases is open to indexer error in their selection of keywords for indexing;
- some databases only give the citation, some an abstract and very few full text. They also vary greatly in their coverage of both published and unpublished grey literature;
- not all databases use a thesaurus—a list of controlled vocabulary used to standardise search terms. PsycLIT has the most sophisticated thesaurus, which is arranged hierarchically, allowing searching for more general or specific terms, as well as for synonyms and related terms. ERIC has a list of terms without any subheadings, and International ERIC does not, as yet, have a thesaurus;
- terms in common usage can be difficult to find, and existing controlled vocabularies may not reflect the latest concepts in a particular field;
- because of the different levels of database sophistication, the same search terms cannot be used consistently.

The literature search identified 265 articles, books, chapters and conference papers, of which 48 fell within the selection criteria for review outlined earlier.

#### *Reviewing the Studies*

The methodology used for reviewing the remaining studies was that developed by the Centre for the Evaluation of Health Promotion and Social Interventions (EPI-Centre) for reviewing evaluations of health promotion interventions. The development of the methodology is described in some detail in Peersman *et al.* (1999). It consists of the following stages:

(i) *Classification of studies and application of inclusion and exclusion criteria.* All the studies found were entered into a Pro-Cite database. Reasons for the exclusion of studies were also logged. Some 96 studies were within the scope of the review, but 48 of these were excluded because they were overviews of the field, or because they did not directly report research findings. However, these studies will inform the background conceptual grounding for the presentation of the findings of the reviewed studies. Of the remaining 48 studies, a further 13 were excluded because, on closer examination, they were also outside the scope of the review.

(ii) *Data extraction.* The remaining 33 studies were reviewed using EPI-Centre software. This consists of a relational database which requires information about each study to be entered and coded. The key descriptors for each study are:

- *how the study can be identified* (citation details, language, the search strategy that

identified it, whether it is published or unpublished, and which keywords are attached to it;

- *support for the study* (funding source, contact person);
- *type of study* (whether it is a process or an outcome evaluation [or both] and whether it is a prospective or retrospective study);
- *a description of the intervention* (including country where it was carried out, the topic area of the intervention, the content of the intervention, the aims, the dates when it was carried out, the theoretical model underpinning it, the intervention site, its length, the type of intervention and the medium of delivery, and who was providing it);
- *a description of the study population* (age, sex, social class, ethnicity etc.) and how they were recruited into the study;
- *planning and process measures* (how the intervention was developed and by whom, how the evaluation was designed and by whom) and to whom the findings were disseminated; and
- *quality of the outcome evaluation* (study design, sample, measurement tools, data analysis method, outcome data, replicability).

(iii) *Quality assessment*. The final judgement to be made is whether the outcome evaluation is 'sound' according to four criteria:

- it has an equivalent control or comparison group, not necessarily randomised;
- it reports pre-intervention data for all individuals/groups (except if the design is the Solomon Four Group, where there are two groups each of experimental and control and where one experimental and one control group does not receive pre-testing);
- it reports post-intervention data for all individuals/groups; and
- it reports on all outcome measures as described in the aims of the study.

As can be seen, these are stringent criteria and reflect a model of research more familiar in medicine than in education. Of the 33 studies reviewed using this method, 11 were found to be sound, using these criteria. A number of the studies used an ABAB design, where an intervention is given and then withdrawn and then reinstated, so that individuals or groups are used as their own controls. This seems to be a common design with psychological research, and avoids some of the difficulties of providing a control group in an educational setting, where it is difficult to provide a self-contained environment in order to be able to isolate the effects of an intervention. As a result of the experience with the EBD review, the EPPI-Centre guidelines have been changed to provide criteria for assessing the quality of these designs also.

Each study was reviewed by two reviewers, and then the results of the separate reviews were compared and any differences discussed and resolved. This process meant that, for each study, between 3 and 5 person-hours was spent. Thus, it can be seen that systematic, rigorous reviews, involving exhaustive searching for studies and a close scrutiny of each study identified, is time-consuming and merits the status of research in its own right.

The next stage of the review will be to analyse the results of the reviewed studies on various dimensions (for example, type of intervention, type of evaluation, population studied, measures of effectiveness) to be able to present some findings about the effectiveness of the different classroom strategies studied.

### Reflections on the Process of the Review and the Adaptation of a 'Medical Model' to Education Research

The choice of EBD interventions as a subject for this review was fortuitous, in that special education has often been accused of using a 'medical model' to locate the problems that children have as 'within child' rather than as a product of the interaction between a child and his or her learning environment (Goacher *et al.*, 1988; Norwich, 1990). We encountered problems of definition (i.e. what do emotional and behavioural difficulties consist of, and what level of severity of difficulty merited the label) and problems about deciding where the boundary between general disciplinary interventions and EBD interventions lay. However, the application of the inclusion criteria for studies kept us focused and helped us to make these decisions.

The inclusion of studies which had an explicit research component (rather than those which described an intervention and gave some generalised commentary about its impact) led to a bias towards studies carried out in the USA, where the tradition of experimental social research is more long-standing, and towards studies carried out within a psychological rather than a sociological perspective. The focus on classroom interventions led us to exclude studies which had a 'whole school' or systems perspective, or which focused on teacher attitudes, rather than on pupil behaviour. Thus, we looked at quite a narrow range of the available studies on support for pupils with EBD.

These may be seen by some to be fatal flaws, but this specificity should enable us to move from general statements about supporting children with EBD to say something more precise and targeted about specific interventions in the classroom and their effectiveness.

The exercise has also demonstrated the lack of available research into strategies to support children with EBD, many of which (for example, Assertive Discipline or Circle Time) are currently being adopted in English schools. Many of the reports excluded from the study were from teachers or psychologists who had adopted certain strategies with a group of pupils and who were reporting in general terms that they appeared to be effective, and were recommending their adoption by others. However, these reports did not include comparison groups and so one could not identify whether it was the intervention *per se* or some other factor which had led to the change in behaviour. Much of this is a matter of the style of reporting the results of evaluations of interventions. Insufficient detail about the aims of interventions, the design of evaluations and the measures used to assess outcomes means that readers of the reports cannot assess their reliability or validity.

Given the dearth of robust evidence in this one small area of educational research, which, more than many, lends itself to experimentation, how can educational research move forward in providing the kinds of evidence about 'what works' which is being called for? One way might be to adopt what Slavin (1986, 1995) has termed 'best evidence synthesis'. Slavin uses a judicial model to argue that, as in some court cases, evidence that would be seen as essential in one case would be disregarded in another case, because better evidence was available, so in reviews of research, other designs than RCTs could be accepted if there were no good RCTs. Nevertheless, he is still arguing for some form of experimental design, with matched controls and attempts to eliminate bias. He is not arguing that descriptive case studies, cross-sectional surveys or ethnographic research designs should be included. However, he does suggest that:

A best evidence synthesis ... should use the evidence at hand to answer

important questions about effects of various treatments, possible conditioning or mediating variables and so on. (Slavin, 1986, p. 10)

This could be interpreted to mean that evidence from case studies and studies using qualitative methodologies could be included in a systematic review to help to illuminate *why* particular interventions are effective (i.e. the process issues or the reasons why particular programme or participant characteristics seem to have an effect on outcomes). Slavin argues that the 'germaneness' of the intervention to the outcome being measured is a key issue and that well-designed RCTs which are not measuring the relevant outcomes are not good evidence. He also suggests that RCTs conducted under highly artificial laboratory conditions do not provide good evidence for what may work in a classroom situation. Thus, there must be a trade-off between authenticity and controlled experimentation. The EPI-Centre (now EPPI-Centre) reflects this concept of 'fitness for purpose', in that both process and outcome studies can be included in a review, and well-designed studies of various types are used as evidence.

The process of undertaking this systematic review of interventions for children with EBD has provided a number of key learning points for those of us involved. Firstly, it has demonstrated that such reviews should be given the status and time allocation of a significant research project. The EPI-Centre has found that reviews in health promotion, including the process of searching for studies, take the time of two full-time research officers over a period of 6–8 months (Harden, personal communication). There is also a considerable time allocation needed by a skilled librarian or information scientist working with researchers to track down the full range of published and unpublished studies.

Secondly, the review should focus on a specific question, and have clearly defined criteria for including and excluding studies. The advice of an expert panel is useful for helping to clarify the scope of a review.

Thirdly, it is important to have clear criteria for assessing the methodological quality of the studies in order to evaluate their 'soundness' and the weight which can be given to their findings. Sound studies do not necessarily only include RCTs, but if one is wanting to say something about the effects of a particular intervention, one would have to be able to judge that against the effects of doing nothing, or of doing something different. Other studies, which might have an explanatory or theoretical role, can be included in a 'best evidence synthesis' to provide possible explanations for the effects or lack of effects of interventions.

### **Implications for Educational Research and Reviews of Research**

The emphasis in current government thinking on 'evidence-based policy and practice' appears to have the potential to limit the diversity of research and review models adopted. The model of research which best lends itself to reviews of the effectiveness of interventions (or 'what works') is some form of experimentation (Oakley, 1998b). It is the case, therefore, that this emerging movement of 'evidence-based policy and practice' in the public sector will steer educational research in the direction of a 'medical model'. The limitations of this approach in terms of the questions which can be addressed need to be debated fully by all those involved. Such an approach will tend to reduce research questions to the pragmatics of technical efficiency and effectiveness. It will not encourage research which explores the wider social, philosophical or ethical issues which are implicit in all social policy decisions.

Experimental research can only answer a fairly limited range of questions, and is not always sensitive to broad questions of values and ethics. David Blunkett confirmed that there will still be a place for critical and 'blue skies' research, and that which addresses the 'big' questions about the social context in which social interventions take place. It remains to be seen whether this commitment by the Government to the broader research agenda is realised in practice.

In terms of the implications for the practice of research, there are some issues which researchers need to confront about the ways in which research into the effectiveness of classroom practices is carried out and reported. Our experience with the EBD review has identified a lack of useful research on classroom interventions and demonstrated that much of what is offered to children is based on teachers' 'craft knowledge'. This is not to be critical of bottom-up strategies, developed by practitioners, which are a key stimulus for change and innovation. However, these developments need to be assessed through rigorous research in order that they can be used effectively as evidence to change practice and improve outcomes for children. The experience of health care has indicated that it is possible to move from a situation where practice is based on tradition or personal preference to one where it is based on sound research evidence. This may be an uncomfortable transition for practitioners and researchers, and one which will need changes in the ways in which researchers and practitioners relate to each other. But it is one which, on the limited evidence of carrying out this review of research, will lead to a much clearer specification by the research and policy communities of the purposes and uses of educational research.

*Correspondence:* Jennifer Evans, Management Development Centre, Institute of Education, 55–59 Gordon Square, London WC1H 0NU, UK; E-mail: j.evans@ioe.ac.uk

## REFERENCES

- ANTMAN, E. M., LAU, J., KUPELNICK, B., MOSTELLER, E. & CHALMERS, T. C. (1992) A comparison of results of meta analyses of randomized controlled trials and recommendations of clinical experts. Treatments for myocardial infarction, *Journal of the American Medical Association*, 268, pp. 240–248.
- BASSEY, M. (2000) Reviews of educational research, *Research Intelligence*, 71, pp. 22–29.
- BLUNKETT, D. (2000) Influence or irrelevance: can social science improve government? *Research Intelligence*, 71, pp. 12–21.
- COE, R. & FITZGIBBON, C. (1998) School effectiveness research: criticisms and recommendations, *Oxford Review of Education*, 24, pp. 421–438.
- COOPER, H. (1998) *Synthesizing Research: a guide for literature reviews*, 3rd edn (London, Sage).
- COOPER, H. & HEDGES, L. (Eds) (1994) *The Handbook of Research Synthesis* (New York, Sage).
- DAVIES, H. T. O., NUTLEY, S. M. & SMITH, P. C. (Eds) (2000) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol, Policy Press).
- DAVIES, P. (1999) What is evidence-based education? *British Journal of Educational Studies*, 47, pp. 108–121.
- DEPARTMENT FOR EDUCATION AND EMPLOYMENT (DfEE) (2000) *Blunkett Rejects Anti-Intellectualism and Welcomes Sound Ideas*, DfEE News 43/00 (London, DfEE).
- DEPARTMENT OF HEALTH (DoH) (1991) *Research for Health: a research and development strategy for the NHS* (London, DoH).
- EVERTON, T., GALTON, M. & PELL, T. (2000) Teachers' perspectives on educational research: knowledge and context, *Journal of Education for Teaching*, 26, pp. 167–182.
- FOSTER, P. (1997) A need to move in mysterious ways, *Times Educational Supplement*, 21 November, p. 14.
- FOSTER, P. & HAMMERSLEY, M. (1998) A review of reviews: structure and function in reviews of educational research, *British Educational Research Journal*, 24, pp. 609–633.
- GILLBORN, D. & GIPPS, C. (1996) *Recent Research on the Achievements of Ethnic Minority Pupils OFSTED Reviews of Research Series* (London, HMSO).

- GIPPS, C. (1997) The funding and management of research in education, *Research Intelligence*, 62, pp. 2–6.
- GLASS, G. (1976) Primary, secondary and meta-analysis of research, *Educational Researcher*, 5, pp. 3–8.
- GLASS, G., MCGRAW, B., SMITH, M. L. & FILBY, N. (1981) *Meta-analysis in Social Research* (Beverly Hills, CA, Sage).
- GOACHER, B., EVANS, J., WELTON, J. & WEDELL, K. (1988) *Policy and Provision for Special Educational Needs* (London, Cassell).
- HALLAM, S. & COWAN, R. (1998) *Is Homework Important for Increasing Educational Attainment?* (London, University of London, Institute of Education).
- HAMMERSLEY, M. (1997) Educational research and teaching: a response to David Hargreaves's TTA lecture, *British Educational Research Journal*, 23, pp. 141–161.
- HARGREAVES, D. H. (1996) *Teaching as a research-based profession: possibilities and prospects (Teacher Training Agency Annual Lecture 1996)* (London, TTA).
- HART, P. & NOLAN, K. (1999) A critical analysis of research in environmental education, *Studies in Science Education*, 34, pp. 1–67.
- MACDONALD, G. (2000) Social care: rhetoric and reality, in: H. T. O. DAVIES, S. M. NUTLEY & P. C. SMITH (Eds) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol, Policy Press).
- MCINTYRE, D. (1997) The profession of educational research, *British Educational Research Journal*, 23, pp. 127–140.
- MORTIMORE, P. & WHITTY, G. (1997) *Can School Improvement Overcome the Effects of Disadvantage?* (London, University of London, Institute of Education).
- NHS CENTRE FOR REVIEWS AND DISSEMINATION (1996) *Undertaking Systematic Reviews of Research on Effectiveness: CRD guidelines for those carrying out or commissioning reviews*, CRD Report No. 4 (York, University of York, NHS Centre for Reviews and Dissemination).
- NOBLITT, G. W. & HARE, R. D. (1988) *Meta-ethnography: synthesizing qualitative data-studies* (Beverly Hills, CA, Sage).
- NORTH THAMES REGIONAL OFFICE OF THE NHS EXECUTIVE (1998) *Leaflet Describing Short Courses in Evidence-Based Health Care* (London, North Thames Regional Office of the NHS Executive).
- NORWICH, B. (1990) *Reappraising Special Needs Education* (London, Cassell).
- NUTLEY, S. & WEBB, J. (2000) Evidence and the policy process, in: H. T. O. DAVIES, S. M. NUTLEY & P. C. SMITH (Eds) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol, Policy Press).
- OAKLEY, A. (1998a) Experimentation and social interventions: a forgotten but important history, *British Medical Journal*, 31, pp. 1239–1242.
- OAKLEY, A. (1998b) Experimentation in social science: the case of health promotion, *Social Sciences in Health*, 4, pp. 73–89.
- OAKLEY, A. (2000) *Experiments in Knowing: gender and method in the social sciences* (Cambridge, Polity Press).
- PEERSMAN, G. V., OAKLEY, A. R. & OLIVER, S. R. (1999) Evidence-based health promotion? Some methodological challenges, *International Journal of Health Promotion and Education*, 37, pp. 59–64.
- PRING, R. (2000) Editorial: educational research, *British Journal of Educational Studies*, 48, pp. 1–10.
- RUDDOCK, J. & MCINTYRE, D. (Eds) (1998) *Challenges for Educational Research* (London, Paul Chapman).
- SACKETT, D. L., RICHARDSON, W. S., ROSENBERG, W. M. & HAYNES, R. B. (1997) *Evidence-based Medicine: how to practice and teach EBM* (Edinburgh, Churchill Livingstone).
- SHELDON, T. & CHALMERS, I. (1994) The UK Cochrane Centre and the NHS Centre for Reviews and Dissemination: respective roles within the Information Systems Strategy of the NHS R & D Programme, coordination and principles underlying collaboration, *Health Economics*, 3, pp. 201–203.
- SLAVIN, R. E. (1986) Best-evidence synthesis: an alternative to meta-analytic and traditional reviews, *Educational Researcher*, 15, pp. 5–11.
- SLAVIN, R. E. (1995) Best evidence synthesis: an intelligent alternative to meta-analysis, *Journal of Clinical Epidemiology*, 48, pp. 9–18.
- TOOLEY, J. & DARBY, D. (1998) *Educational Research: a critique. A Survey of Educational Research* (London, Office for Standards in Education).
- WEISS, C. H. (1979) The many meanings of research utilisation, *Public Administration Review*, 39, pp. 426–431, cited in: S. NUTLEY & J. WEBB (2000) Evidence and the policy process, in: H. T. O. DAVIES, S. M. NUTLEY & P. C. SMITH (Eds) *What Works? Evidence-based Policy and Practice in Public Services* (Bristol, Policy Press).