Carfax Publishing
Taylor & Francis Group

Research Evidence, Knowledge Management and Educational Practice: early lessons from a systematic approach

ANN OAKLEY

EPPI-Centre and Social Science Research Unit, Institute of Education, University of London, UK

ABSTRACT   'Evidence-informed' education is part of a general move internationally to found policy and practice on a more reliable scientific base. This paper describes a five-year initiative in this area funded by the English Department for Education and Skills in 2000. The paper describes the background to the initiative, the infrastructure developed within it for supporting research synthesis, and some of the challenges experienced by the first review groups.

Introduction

Education systems can be under pressure for many reasons, but an important challenge today is the move towards a 'knowledge-based' society. A recent OECD report drew an analogy between this transformation and the earlier one during the industrial revolution, when machinery was first developed to produce machinery. The fundamental pre-requisite for a full-scale move to a knowledge-based society is the development of systematic methods for applying knowledge to the *production* of knowledge (OECD, 2000, p. 12). Implied in this is the need for a systematic approach also to the management and dissemination of knowledge, including especially the interface between knowledge and practice.

This paper discusses a particular approach to knowledge management and educational practice within the context of the evolution of the 'evidence-informed' education movement in England. This movement is part of a general awakening of interest across Europe and North America in the possibilities of basing public policy on a reliable and user-friendly evidence base. The paper takes a synoptic look at the background and context in which this movement developed, gives a descriptive account of the setting up of an initiative sponsored by the English Department for Education and Employment (now the Department for Education and Skills) to promote the synthesis of research evidence, and discusses some of the early challenges experienced in taking this initiative forward.

Swings and Roundabouts: the 'science' of educational knowledge

The National Research Council in the USA has described the accumulation of scientific knowledge over time as 'circuitous and indirect', commenting that this process '. . . often traverses highly contested territory . . .' (Shavelson & Towne, 2001). The debate about whether education is an art or a science has gone on for a long time (Rusk, 1919). Central to this debate is an argument about the methods used to generate knowledge about effective and appropriate educational practices: theories about education, multiple understandings about the role and purpose of education, and teachers' individual 'expert' experiences, on

the one hand, or systematically accumulated, replicable and generalisable evidence of the kind used to judge the therapeutic status of professional interventions in other fields, on the other. The other historical constancy is a degree of consensus about the chaotic and methodologically impoverished nature of much educational research (Eisner, 1979; Gage, 1972; OECD, 2000). 'Expert opinions, pooled judgements, brilliant intuitions, and shrewd hunches are frequently misleading,' wrote Julian Stanley in 1957. 'Ultimately they must be tested by the careful gathering of evaluative data if education is to advance on the basis of sound principles' (Stanley, 1957, p. 198). 'Given the huge amounts of educational research conducted over the last fifty years or more,' remarked David Hargreaves in his lecture to the Teacher Training Agency in 1996, 'there are few areas which have yielded a corpus of research evidence regarded as scientifically sound and as a worthwhile resource to guide professional action' (Hargreaves, 1996, p. 2).

In the late 1970s and early 1980s a group of health service researchers in Oxford prepared the ground for evidence-based medicine by beginning a programme of systematic reviews on the effectiveness of health care interventions.

The Cochrane Collaboration (http://www/cochrane.org) is now an international network of researchers, academics, practitioners and users committed to the principles of managing knowledge in such a way that it is quality assured, accessible and cumulative. Cochrane reviews mostly, but not exclusively, focus on randomised experimental studies, and include a number of relevance to education, for example interprofessional education, and school-based driver education for the prevention of traffic crashes. The Cochrane Collaboration's sister organisation, the Campbell Collaboration (http://www.campbell collaboration.org), is adapting Cochrane methodology to prepare, maintain and disseminate systematic reviews of social interventions, but, being younger, is still working out its methods. The climate of thinking about public policy both in the USA and in the UK has shifted more firmly into the 'evidence' mode, with policy-makers declaring the need for a sounder science on which to base more 'joined up' decisions about policy issues (Cabinet Office, 1999). It has also become clear that there is a considerable forgotten history of scientific policy experimentation and research synthesis (Oakley, 1998a). A return to this tradition is urgently needed, not just for policy-makers or even to improve education and other professional practices: it is the ordinary citizen who is potentially most disadvantaged by the lack of an open, systematic basis of evidence concerning the many interventions that intrude into every corner of life (Smith, 1996).

These developments make it more difficult to hide failures in a field of research activity which is of such critical public importance as education. Unlike health care, education is a compulsory intervention. Like health care professionals, teachers believe that they are only acting in the best interests of those at the receiving end of their work, and that it is their unique vocation to 'know' the best forms of practice. These parallels between education and health care have driven some recent criticisms of educational research, and have particularly drawn attention to the over-reliance in much educational practice on procedural 'craft' knowledge rather than more open research-based evidence (Hargreaves, 1996; Hillage *et al.*, 1998; McIntyre & McIntyre, 1999; Tooley & Darby, 1998). Another way to put this is to describe the educational sector as dominated by non-codified but potentially codifiable knowledge (OECD, 2000, p. 19). There are areas such as HIV/AIDS education where the health care and education systems have a shared responsibility to introduce a greater degree of codification into the knowledge base. The disadvantages of *not* doing this can, in such circumstances, literally be fatal. A recent project helping policy-makers, practitioners and researchers in southern Africa to develop a more scientific approach to HIV/AIDS education demonstrated how much of an unmet need there can be on the ground for support in moving towards a more systematic evidence-base (Stewart *et al.*, 2001).

What is Evidence?

Davies (1999, p. 109) defines 'evidence-based education' as operating at two levels: the *use* of educational research and the *establishment* of sound research evidence through academic practices which attend to the criteria of scientific validity, high quality and practical relevance. The National Research Council in the USA notes six fundamental scientific principles that apply in the field of educational research: posing significant questions that can be answered empirically; linking research to relevant theory; using methods that permit direct investigation of the question; providing a coherent and explicit chain of reasoning; replicating and generalising across studies; and disclosing research to professional scrutiny and critique (Shavelson & Towne, 2001). The underlying argument is that much research can have practical application, but at present the potential that research has to inform policy and practice is not being realised. Two main reasons for this are the reluctance of the teaching profession to view research as relevant, and the tendency of the academic community to undertake and disseminate research in a largely inward-looking and self-seeking way (Hargreaves, 1996).

Systematic reviews are the primary method for managing knowledge in the evidence movement approach. This is because they synthesise the findings of many different research studies in a way which is explicit, transparent, replicable, accountable and (potentially) updateable. In these respects, systematic reviews differ from traditional literature reviews, which commonly focus on the *range* and *diversity* of primary research using a selective, opportunistic and discursive approach to identifying and interpreting relevant literature (Badger *et al.*, 2000; Davies, 2000). In traditional 'narrative' reviews, there is often no clear audit trail from primary research to the conclusions of the review, and important research may be missing, resulting in biased and misleading findings, and leading to puzzling discrepancies between the findings of different reviews.

Although there is a strong tradition of non-systematic reviews in social science, there are usually more of the systematic kind than might be expected. For example, a survey of effectiveness reviews in health promotion identified 398 reviews, of which 75 could be considered systematic (Peersman *et al.*, 1999). There are over 300 systematic reviews in the Campbell Collaboration's Social, Psychological, Educational and Criminological Trials Register (Petrosino *et al.*, 2000).

The EPPI-Centre

Recognising the importance of systematic reviews in improving the evidence available to educational practitioners and policy-makers, the English Department for Education and Skills funded the Evidence for Policy and Practice Information and Coordinating Centre at the Social Science Research Unit, the Institute of Education, University of London, in 2000 to undertake a five-year programme of work in this area. The main purposes of the funding are: to support groups of researchers, practitioners and users external to the EPPI-Centre—Review Groups (RGs)—to carry out systematic reviews; to develop the tools, procedures and training needed to facilitate review group work; and create and sustain accessible web-based databases of systematically keyworded research literature, data from primary studies, and the results of systematic reviews. The EPPI-Centre initiative is focusing mainly on schools and students aged 0–16 years, although one review (not funded by the DfES) is concerned with the post-compulsory phase.

The EPPI-Centre's DfES work is about *managing* rather than *generating* knowledge. It is based on a number of key principles about systematic reviews. Most of these are common to the practice of systematic reviews in general (see Chalmers & Altman, 1995; NHS Centre, 2001; Light & Pillemer, 1984), although some flow from the EPPI-Centre's own previous experience of conducting systematic reviews of social interventions in areas such as health

promotion and sex education (see Harden *et al.*, 2001; Shepherd *et al.*, 2001). In the EPPI-Centre approach, a systematic review:

- involves a series of explicit, discrete and standard stages;
- means specifying a particular, answerable research question, and criteria about what kinds of studies (by topic/population group/setting/research design etc.) will be included in, and excluded from, the domain of literature to be surveyed;
- requires clarity about which literatures will be searched for relevant studies, and how;
- includes making explicit, justifiable decisions about the methodological quality of studies regarded as generating reliable findings;
- needs some method of integrating the findings of individual, good quality studies; and
- is credible only if it has involved input from research users at all stages of the review process.

These principles are laid out and given operational flesh in a manual for RGs available as an online tool (http://eppi.ioe.ac.uk). Also available on the same website is a document outlining a keywording strategy for classifying educational research studies, and references to some of the EPPI-Centre's own systematic reviews which develop traditional review methodology in order better to meet the challenges of systematic reviews in social science. For example, a criterion for including studies in a review of workplace health promotion interventions was that interventions needed to have been developed in response to needs or views expressed by employees (Harden *et al.*, 1999a). A review of peer-delivered health promotion for young people included 'qualitative' studies of the processes involved in implementing different programmes (Harden *et al.*, 2001). And a recently completed series of reviews on barriers to, and facilitators of, young people's mental health, healthy eating and physical activity set research on young people's views alongside evaluations of health promotion interventions to see what could be learnt about matches and mismatches between the policy and practice conclusions emerging from each type of research (Shepherd *et al.*, 2001). It is clear from these examples, and from the steps and procedures set out in the EPPI-Centre training materials, that systematic reviews can include a wide range of study designs.

Ten RGs have been established in the first two years of the EPPI-Centre education initiative in the following areas: English teaching; assessment and learning; school leadership; gender and education; post-compulsory education (funded by the Learning and Skills Development Agency); inclusive education; early years; thinking skills approaches to effective learning and teaching; modern languages; and the impact of continuing professional development on classroom teaching and learning. Applications from a 'third wave' of RGs are currently being considered. The methods for facilitating the review work of these groups were piloted in a systematic review of strategies to support pupils with emotional and behavioural difficulties (EBD) in mainstream primary classrooms conducted by two researchers at the Institute of Education and the National Foundation for Educational Research (Evans & Benefield, 2001).

EPPI-Centre RGs go through a formal registration process which includes peer refereeing of plans and protocols. Each review undertaken by an RG requires a detailed review protocol which defines the review question, inclusion and exclusion criteria and strategies for searching the literature. All protocols are placed on the EPPI-Centre website for open comment. Reviews may often be done in two stages: *a mapping stage*, in which relevant literature is captured and systematically keyworded to provide a descriptive account of the research effort in that particular area; and *an in-depth review stage*, in which a subset of the literature (of particular interest/relevance to the review question) is examined and interrogated in more detail and data extracted from primary studies. An important feature of the reviewing software developed by the EPPI-Centre is that it allows data to be entered on a range of study designs. These data are then analysed together for individual reviews, but

they can be combined differently to answer other review questions; reviews can also easily be updated by adding data from new primary studies.

Each RG is scheduled to produce one review per year. The completed reviews are available online free of charge in a database called REEL (Research Evidence in Education Library), which was launched in summer 2002.

The First Reviews: research questions and study designs

The first reviews from the six RGs registered in the first year show considerable diversity in terms of the questions asked and the kinds of studies included in the review. The review questions are: 'what is the impact of networked ICT on literacy learning in English for pupils aged 5 to 16?' (the English RG); 'what is the impact of summative assessment on pupils' motivation for learning?' (the Assessment RG); 'what is the impact of school leadership and management on student/pupil outcomes?' (the Leadership RG); 'what kind of strategies improve equal opportunities in gender for pupils in mixed sex primary schools?' (the Gender RG); 'what is the impact of the financial circumstances of learners on their engagement with learning?' (the Post-compulsory RG); and, 'what evidence is there that mainstream schools can act in ways which enable them to respond to diversity so as to facilitate participation by all students in the cultures, curricula and communities of those schools?' (the Inclusive Education RG).

Types of research given prominence in the review protocols are as follows: outcome evaluations, process evaluations and descriptive studies (the English RG); intervention and non-intervention studies (the Assessment RG); intervention studies and school effects and improvement research (the Leadership RG); outcome evaluations and process evaluations (the Gender RG); studies 'which seek to understand the links between financial circumstances and engagement' (the Post-compulsory RG); ethnographic case studies, outcome and process evaluations, survey research and other types of case study (the Inclusive Education RG). Table 1 shows the classification of types of studies finally included in the in-depth stages of the first four reviews (Assessment, English, Gender and Inclusive Education) to appear on REEL. Of the 71 studies examined in depth in the four reviews, 35% (25) evaluated the outcomes of some type of educational intervention, a further 37% (26) were descriptive studies, and the remainder (28%–20) reported either evaluations of process or a mixture of process and outcome evaluations.

Challenges of the Review Process

The process of managing knowledge about educational practices through central support for RG work described in this paper is itself by way of being an experiment. Although there is

Table 1 **Types of studies included in the in-depth stages of the first four EPPI-Centre systematic reviews**

Review	Outcome evaluation	Process & outcome evaluation	Process evaluation	Descriptive study	Total
Assessment	13	1	3	2	19
English	8	–	8[a]	–	16
Gender	1	2	1	5	9
Inclusion	3	–	5	19	27
Total	25 (35%)	3 (4%)	17 (24%)	26 (37%)	71 (100%)

[a] 'process evaluations and other studies'

increasing interest in many circles in systematic reviewing as an essential part of the move towards a knowledge-based society, there is, so far as we are aware, no directly comparable model in which support, guidance, resources and 'quality control' are provided to groups of external reviewers by a central organisation. It will be some time before the success or otherwise of the EPPI-Centre enterprise, in terms of impact on policy or practice decisions, can be judged. But some lessons have emerged in the first two years' experiences of this approach to managing educational research knowledge. The discussion below draws on the pilot EBD review, RGs' reports of their first year's work, observations of the training workshops and other EPPI-Centre contacts with RGs, and the first four reviews published on REEL.

The challenges can be divided into: (a) political, (b) technical, and (c) conceptual/intellectual.

Political Challenges

The venture into the area of systematic reviews of educational research described in this paper is taking place in a highly contested domain. There is a sizeable voice in the educational research world objecting to what is seen as the unwarranted transfer to the educational domain of a 'positivistic' model of knowledge dominated by a 'hierarchy of evidence', limited questions about 'what works', outdated notions about the role of 'procedural objectivity' and a disregard for the tenets of postmodernism which question the validity of any *a priori* framework of enquiry (see Atkinson, 2000; Elliott, 2001; Hammersley, 2001). There are accusations that a coordinated structure for preparing systematic reviews in education reeks of unwarranted monitoring and control, and represents an attempt by government to dictate the form educational research should take, if not the actual topics of enquiry (see Ball, 2001). In some quarters, the whole exercise is perceived as a waste of time because there are few 'golden insights', and nothing obviously newsworthy about the conclusions of reviews which report such a paucity of reliable literature (Budge, 2002).

What could loosely be termed 'the anti-evidence movement' in educational research sets itself against the whole business of systematic research synthesis on a number of grounds: that it incorrectly presupposes a rational process of evidence-informed policy and practice; is limited to certain kinds of quantitative studies; ignores the importance of theoretical development; exhibits a limited, and limiting, emphasis on 'what works?' questions; dismisses the 'craft' knowledge of teachers as unimportant, and mistakenly assumes that education shares with medicine a common epistemological and professional base (Atkinson, 2000; Ball, 2001; Hammersley, 1997; Hargreaves, 1997; Hulme, 2002). Like randomised controlled trials with which it is often semantically confused, systematic research synthesis may be seen as part of a trend towards 'instrumentalism' which puts 'nails in the coffin of academic freedom' (Atkinson, 2000, p. 318).

Interestingly, these debates repeat themes of earlier attacks on evidence-based medicine and evidence-based health promotion (Oakley, 1998b), and they ignore a reputable history of research synthesis methods being developed precisely as tools for assessing social interventions in people's lives, such as education (Glass *et al.*, 1981). It is not the purpose of this paper to discuss in detail the 'anti-evidence' position in education (see Elbourne & Gough, 2002; Gough & Elbourne, 2002; Oakley, 2001; Oakley, in press). Rather, the point is to acknowledge that those engaged in doing systematic reviews in education must often actively contest with their colleagues and others the position that this approach is a suspicious and inappropriately mechanistic procedure for reaching conclusions about what educational research can reliably tell us. Such antagonism and misunderstanding multiplies the difficulties of work in an area that is challenging for other structural and intellectual reasons.

Technical Challenges

The technical challenges of systematic reviews relate to skills and training, the limitations of electronic databases, low yields of usable studies, and time and cost implications.

Systematic reviewing is an activity that calls for particular kinds of skills and resources. These are not necessarily part of academic or practitioner training, or of the normal contexts in which academics and practitioners work. Indeed, very little of the research training currently on offer in the social sciences or in teacher education includes components which prepare people for systematic review work. A primary challenge has therefore been the *lack of experience of the skills and procedures* (e.g. searching, classifying literature, keywording, data extraction) needed in systematic reviewing. This is a problem for the core academic members of an RG, but also for the involvement of teachers and other educational research users.

A second issue, related to the relative novelty of systematic reviewing as an educational research activity, is the *reliance on relatively 'undeveloped' electronic databases* which are sources of the primary studies to be included in reviews. Many of these databases (e.g. ERIC and PsycLIT) lack comprehensive keywording, any thesaurus of standardised search terms or sophisticated search strategies, and each requires individualised approaches. It can take a long time to work out what search terms to use for each database in order to identify literature for a particular review. An iterative process can be useful; in this, the productivity of different search strategies is tested against a list of known studies at the end of the review process (see Harden *et al.*, 1999b). These difficulties with electronic databases highlight the importance of personal contacts and of handsearching—both time-consuming exercises—as sources of relevant material. For example, in its first review the Assessment and Learning RG spent a frustrating four months on searching 63 journals electronically and by hand with a much greater use of handsearching than initially anticipated, largely because of the inadequacies of electronic databases.

A third challenge is *the relatively low yield of usable studies* derived from the searching process. Table 2 shows the searching process for the four reviews currently on REEL. Only 0.3% of the initial citations were sufficiently relevant to be reviewed in depth in the two reviews (English and Inclusive Education) which give a figure for initial citations. Of the studies that met the inclusion criteria on title and abstract in the four reviews, 6% were reviewed in detail. This experience of having to search haystacks to find needles is common in systematic reviews (Oakley & Fullerton, 1996), and it can consume substantial resources in time and labour.

Table 2 **Finding reliable evidence: the first four EPPI-Centre systematic reviews**

Review	Citations found	Met inclusion criteria (title & abstracts)	Met inclusion criteria (full paper)
Assessment	–	183	19
English	1,871	361 (mapping stage)	188 (mapping stage)
			16 (in-depth stage)
Gender	–	497	9
Inclusion	14,692	336	41 (27 studies)

Thus, fourthly, *the length of time systematic reviews take* can be a surprise to people new to systematic reviewing. Not only is searching for literature a lengthy business, but scrutinising titles and abstracts for relevant information is often exhausting and disappointing, retrieval of hard copies of studies for in-depth review is another time-consuming task, and extracting data from individual studies for the review can take 3−5 person hours per study. One implication of this is that 'such reviews should be given the status and time allocation of a significant research project' (Evans & Benefield, 2001, p. 539).

The resource 'burden' of systematic reviewing results in a further challenge: *the significant real cost of producing reviews*. In the case of the EPPI-Centre initiative, the DfES, through the EPPI-Centre, provides some pump-priming money for the RGs (£20k for the first year, £15k thereafter), but experience of the first RGs suggests that the real cost per review is likely to be closer to £75k. The in-house reviews the EPPI-Centre does for other funders run from £50−£80k per review, and detailed reviews which tackle important methodological questions at the same time can cost in excess of £100k. Given these resource demands, pump-priming money can be rapidly used up, resulting in the possibility of less-than-ideal shortcuts in the review process.

Conceptual/Intellectual Challenges

The three main conceptual and intellectual challenges experienced in the first wave of EPPI-Centre reviews concern defining the initial research question; deciding how (or whether) to 'weight' studies in terms of methodological quality; and operationalising the notion of 'user involvement' throughout the review process.

Defining the initial research question for a systematic review can be surprisingly tricky. Groups may find their collective expertise in an area not necessarily an aid to clear thinking about how to identify a single review question and define the key terms included in such a question to the satisfaction of all. For example, the Inclusive Education RG had problems defining the terms 'inclusive' and 'organisational characteristics' in its initial review question in such a way that these were acceptable to the different ideological positions of RG members (Torgerson *et al.*, in press). ('Inclusive' became the promotion of 'participation of all students in the culture, curricula and communities' of their schools, and 'organisational characteristics' became 'ways in which schools can act'.) Further definition may be necessary for the question asked in the in-depth stage of the review. For instance, the Gender RG's in-depth review question narrowed the scope of its initial review question to become 'what kind of strategies delivered by teachers or researchers in the classroom reduce stereotypical gender constructions among girls and boys in UK mixed sex primary schools?' The final protocol for a review can be a lengthy task: the English RG's review protocol went through six drafts, for example. One reason why specifying review questions and criteria can be such a lengthy, iterative process is because such discussions are precisely not just technical; they provide a forum for debate about important theoretical and ideological issues. Systematic reviews are in this sense as much a theory-based as a pragmatic form of inquiry, a point often not fully realised by people before they embark on a systematic review, nor by those who see systematic reviews as a suspiciously atheoretical activity.

Equally, if not more, problematic, is the question of methodological quality. *Deciding how to define 'sound' studies* (of whatever research design) can be a contentious issue. If systematic reviews are not simply to include all the literature found, then there has to be some screening out of empirical studies that do not offer a particularly credible account of how the data were collected and how the findings are based on these. If this is not done, there is a danger that the conclusions of the review may themselves be unsound in directing researchers, practitioners, policy-makers and other users of research evidence towards

policies and practices that will not achieve intended goals. The notion that some research studies may be 'better' methodologically than others can offend the egalitarianism of some systematic reviewers. There can be obvious problems in defining some studies as less 'good' than others when the field is relatively small, and researchers all know one another. Even when the need for quality criteria is agreed, it can be difficult to settle on particular criteria for identifying good quality studies. It can also be very difficult to decide when such criteria have been met, since a surprising proportion of studies are very inexplicit, not to say vague, about the designs they have used.

Table 3 shows the conclusions of the first four reviews on methodological quality. The term 'weight of evidence' refers to guidance offered by the EPPI-Centre on a pragmatic way to take account both of methodological soundness and appropriateness and of relevance when considering how much weight should be placed on the findings of a primary study included in a systematic review. It is suggested that studies be assessed on soundness of method—the extent to which a study was carried out according to accepted best practice within the terms of the method used; on appropriateness of study type to answer the particular research question; and on relevance of the topic focus of the study to the question being posed in the systematic review question. An overall judgement of 'weight of evidence' is then made on the basis of these assessments. Three of the four reviews followed this approach (details of exactly how the judgements are made can be found in the individual reviews). The fourth review, by the Gender and Education RG, did not offer any weight of evidence assessments, because the review's authors disagreed with the 'positivist assumptions' and the suggestion of 'fixed' criteria of quality they perceived as underlying such assessments (section 7.3, paras 7 & 9). However, they stated that they 'required a report of the methods and processes of the intervention itself, adequate to describe sufficiently what had taken place, and a credible account of the way in which study findings were reached' (section 6.2, para 3) in deciding which studies to include in a narrative synthesis. They reported on these characteristics of included studies in Appendix D of the review. As well as these overall judgements of methodological quality made in the four reviews, table 3 shows separately the authors' assessments, which were made in three of the reviews, on the specific issue of soundness of design: how well studies of different types were designed and carried out.

Of the 71 studies reviewed in-depth in the four reviews, a minority—24 (34%)—were assessed as yielding 'high quality' weight of evidence. A further 21 (29%) were considered

Table 3 **Overall weight of evidence, and soundness of study design: studies included in the in-depth stages of the first four EPPI–Centre systematic reviews**

Review	Overall weight of evidence			Soundness of study design[a]			Total
	High	Medium	Low	High	Medium	Low	
Assessment	12	6	1	10	6	3	19
English	2[b]	3	11[c]	–	5[d]	11[e]	16
Gender	4[f]	4[f]	1[f]	–	–	–[g]	9
Inclusion	6	8	13	6	10	11	27

[a] soundness of study within design: how well was it designed and carried out?
[b] 'high/medium' or 'medium/high'
[c] includes 'medium/low' and 'low/medium'
[d] includes one 'medium/high'
[e] includes 'low/medium' and medium/low'
[f] adequate report of the methods of processes of the intervention and credible account of the way in which findings reached
[g] this was not assessed in the Gender review

to have produced 'medium' and 26 (37%) 'low' quality evidence. With respect to the single criterion of study design, 16 of the 62 studies (26%) assessed on this in three reviews were considered to have been well designed and carried out, 25 (40%) were considered to have been poorly designed and implemented, and the rest—21 (34%) fell in-between.

Combining the evidence from tables 2 and 3, it is clear that the yield of 'high' or 'medium' quality studies as a proportion of initial citations in these four reviews was disappointingly low. Evidence considered 'high' or 'medium' quality was offered by 0.1% (19/16,563) of initial citations for the two reviews that gave a figure for the results of their citation searches, and by 3.3% (45/1,377) of studies that met inclusion criteria on title and abstract in all four reviews; these figures were 0.1% (21/16,563) and 4.2% (37/880) respectively for soundness of study design (the latter figure excludes the Gender review, where soundness of study design was not assessed). Thus, not only can assessing methodological quality be a problematic issue for educational research reviewers, but, when quality criteria *are* applied, the conclusion may be that educational research suffers from a paucity of high quality studies. Significantly, this conclusion, in the case of the first EPPI-Centre reviews, was reached by authors who are distinguished academics and leaders in their field. It is a finding which supports a key point about the state of educational research made in recent critiques. Indeed, the first reviews on REEL all have recommendations to make about the need to improve methodology in educational research. For example:

> . . . teachers and fellow researchers need to be able to assess whether the research evidence is reliable; the reporting of at least basic aspects of the research methods should be prioritised as good practice. (Gender Review)

> The answer to the research question for the indepth review . . . is inconclusive . . . because there is insufficient research of high quality . . . we have pointed out shortcomings, as we see them, in many of the studies we have examined . . . (English Review)

> Given the problems with methodological quality noted . . . there is a need for studies which are methodologically sound . . . The lack of detail about methodology in much of the literature suggests that practices of research reporting need to change. (Inclusive Education Review)

A third conceptual/intellectual challenge is that of *operationalising 'user involvement'*, which is a key principle of the EPPI-Centre approach to systematic reviews, following the precept that user-led research synthesis is essential if research evidence is to be transformed into practical research knowledge (Desforges, 2000). There is much uncertainty about what the term 'user' means; the boundaries between 'users' and 'producers' of research can be fluid. It has clearly been easier to date to identify as interested users practising teachers and representatives of teacher organisations than representative parents or students. The four reviews currently on REEL involved teachers and policy-makers more than other user groups in the process of conducting the review. Of the 15 user 'perspectives' on the reviews' findings attached to the four reviews, four are written by teachers, four by policy-makers, two by a parent and a parent school governor, two by other school governors, and one by a student; two others were written by teachers on the basis of discussions with students.

Conclusion

The work described in this paper represents the first major UK resource for systematic review work exclusively within the field of education. This is two years ahead of a similar initiative,

which has just been announced, funded at the level of $18.5 million by the US Department of Education to establish a national 'What Works' Clearinghouse for educational research.

The paper has argued that practical experience in reviewing educational research is an important part of current moves towards a knowledge-based society. The DfES's EPPI-Centre initiative is also an important reference point for contemporary debates about the application of the 'evidence movement approach' to educational research.

The EPPI-Centre approach uses a process of collaborative and open working to systematise knowledge about educational practice. A major endpoint of this process—a freely available updateable library of reviews of research evidence—is already being heavily used. In the first three months REEL has been visited by people in 32 countries, who have downloaded 2,197 copies of the first reviews and 1,112 copies of EPPI-Centre education tools. The work of the Centre and of the RGs over the past four years has been a mutual learning exercise, in which both have learnt from one another. Experience with the first RGs suggests that many criticisms of the attempt to provide, through systematic reviewing, a firm evidence base for educational policy and practice are ill-founded in promoting systematic inquiry as necessarily at odds with methodological pluralism. The finding that systematic searches may reveal little methodologically reliable evidence (however defined) is a familiar lesson from systematic reviews in other areas (Barr *et al.*, 1999; Mytton *et al.*, 2002; Oakley & Fullerton, 1986); whether the field of education is notably more deficient in this respect is a question for further research. Similarly, we should now have the basis on which to explore the important question about the extent to which systematic reviews in education may produce consistently different findings from those of traditional non-systematic reviews.

Many of the anxieties currently being voiced about research synthesis are based on misunderstandings of what it actually in practice means: whose interests it is for, who will produce and use it, and who will benefit from it. There are, however, real problems in carrying out systematic reviews of educational research, relating to particular technical, conceptual and intellectual challenges. The future of systematic reviews in education will depend to a large extent on the commitment of the educational research community to making progress on these fronts.

Correspondence: Professor Ann Oakley, SSRU, Institute of Education, 18 Woburn Square, London, WC1H 0NR, UK.

Acknowledgements

References

ATKINSON, E. (2000) In defence of ideas, or why 'what works' is not enough, *British Journal of Sociology of Education*, 21, 3, 317–330.

BADGER, D., NURSTEN, J., WILLIAMS, P. & WOODWARD, M. (2000) Should all literature reviews be systematic? *Evaluation and Research in Education*, 14, 3&4, 220–230.

BALL, S.J. (2001) 'You've been NERFed!' Dumbing down the academy: National Educational Research Forum: 'a national strategy—consultation paper': a brief and bilious response, *Journal of Education Policy*, 16, 3, 265–268.

BARR, H., HAMMICK, M., KOPPEL, I. & REEVES, S. (1999) Evaluating interprofessional education: two systematic reviews for health and social care, *British Educational Research Journal*, 25, 4, 533–544.

Budge, D. (2002) Good in theory, *Times Educational Supplement*, 12 July.

Cabinet Office (1999) *Professional Policy-making for the 21st Century*, London, Stationery Office.

Chalmers, I. & Altman, D. (eds) (1995) *Systematic Reviews*, London, BMJ Publishing.

Davies, P. (1999) What is evidence-based education? *British Journal of Educational Studies*, 47, 2, 108–121.

Davies, P. (2000) The relevance of systematic reviews to educational policy and practice, *Oxford Review of Education*, 26, 3&4, 365–378.

Desforges, C. (2000) Familiar challenges and new approaches: necessary advances in theory and methods in research on teaching and learning. The Desmond Nuttall/Carfax Memorial Lecture. Cardiff, BERA Annual Conference.

Eisner, E.W. (1979) *The Educational Imagination: on the design and evaluation of school programs*, New York, Macmillan.

Elbourne, D. & Gough, D. (2002) Fit for purpose: appropriate methods for primary research and systematic reviews of research to provide evidence to inform potential users of research in education. Paper given at York-IPN International Symposium, Evaluating Science Curriculum Innovation: how do we know we have made a difference? York, 9–11 May.

Elliott, J. (2001) Making evidence-based practice educational, *British Educational Research Journal*, 27, 5, 555–574.

Evans, J. & Benefield, P. (2001) Systematic reviews of educational research: does the medical model fit? *British Educational Research Journal*, 27, 5, 527–541.

Gage, N.L. (1972) *Teacher Effectiveness and Teacher Education: the search for a scientific basis*, Palo Alto, CA, Pacific Books.

Glass, G.V., McGaw, B. & Smith, M.L. (1981) *Meta-Analysis in Social Research*, Beverly Hill, CA, Sage.

Gough, D. & Elbourne, D. (2002) Systematic research synthesis to inform policy, practice and democratic debate, *Social Policy and Society*, 1, 3, 225–236.

Hammersley, M. (1997) Educational research and teaching: a response to David Hargreaves' TTA lecture, *British Educational Research Journal*, 23, 2, 141–161.

Hammersley, M. (2001) On systematic reviews of research literatures: a 'narrative' response to Evans and Benefield, *British Educational Research Journal*, 27, 5, 543–554.

Harden, A., Oakley, A. & Oliver, S. (2001) Peer-delivered health promotion for young people: a systematic review of different study designs, *Health Education Journal*, 60, 4, 339–353.

Harden, A., Peersman, G., Oliver, S., Mauthner, M. & Oakley, A. (1999a) A systematic review of the effectiveness of health promotion interventions in the workplace, *Occupational Medicine*, 49, 8, 540–548.

Harden, A., Peersman, G. & Oakley, A. (1999b) Identifying primary research on electronic databases to inform decision-making in health promotion: the case of sexual health promotion, *Health Education Journal*, 58, 290–301.

Hargreaves, D. (1996) Teaching as a research-based profession: possibilities and prospects. Teacher Training Agency Annual Lecture, London, TTA.

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

Hargreaves, D. (2001) The Desmond Nuttall Memorial/Carfax Lecture. BERA Annual Conference.

Hillage, J., Pearson, R., Anderson, A. & Tamkin, P. (1998) *Excellence in Research in Schools*, London, Department for Education and Employment/Institute of Employment Studies.

Hulme, B. (2002) Policy transfer in evidence based practice in education. BERA Annual Conference (Abstract), University of Exeter 12–14 September.

McIntyre, D. & McIntyre, A. (1999) *Capacity for Research into teaching and Learning.* Final Report to ESRC Teaching and Learning Programme, Cambridge, School of Education, University of Cambridge.

Mytton, J.A., Digiuseppi, C., Gough, D.A., Taylor, R.S. & Logan, S. (2002) School-based violence prevention programs, *Archives of Pediatrics and Adolescent Medicine*, 15, 752–762.

NHS Centre for Reviews and Dissemination (2001) *Undertaking Systematic Reviews of Research on Effectiveness. Guidelines for those carrying out or commissioning reviews*, York, Centre for Reviews and Dissemination, University of York.

Light, R.J. & Pillemer, D. (1984) *Summing Up: the science of reviewing research*, Cambridge, MA, Harvard University Press.

Oakley, A. (1998a) Public policy experimentation: lessons from America, *Policy Studies*, 19, 2, 93–114.

Oakley, A. (1998b) Experimentation in social science: the case of health promotion, *Social Sciences in Health*, 4, 2, 73–89.

Oakley, A. (2001) Making evidence-based practice education: a rejoinder to John Elliott, *British Educational Research Journal*, 27, 5, 575–576.

Oakley, A. (in press) Social science and evidence-based everything, *Educational Review*.

Oakley, A. & Fullerton, D. (1986) The lamp-post of research: support or illumination? In: A. Oakley & H. Roberts (eds) *Evaluating Social Interventions*, Ilford, Barnardos.

OECD (2000) *Knowledge Management in the Learning Society*, Centre for Educational Research and Innovation.

Oliver, S., Peersman, G., Harden, A. & Oakley, A. (1999) Discrepancies in findings from effectiveness reviews: the case of health promotion for older people in accident and injury prevention, *Health Education Journal*, 58, 66–77.

Peersman, G., Harden, A., Oliver, A. & Oakley, A. (1999) *Effectiveness Reviews in Health Promotion*, London, EPPI-Centre.

Petrosino, A., Boruch, R.F., Rounding, C., McDonald, S. & Chalmers, I. (2000) The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions, *Evaluation and Research in Education*, 14, 3&4, 206–219.

Rusk, R.F. (1919) *Experimental Education*, London, Longmans, Green & Co.

Shavelson, R.J. & Towne, L. (eds) (2001) *Scientific Inquiry in Education*, Washington DC, National Academy Press.

Shepherd, J., Garcia, J., Oliver, S., Harden, A., Rees, R., Brunton, G. & Oakley, A. (2001) *Barriers to, and Facilitators of, the Health of Young People: a systematic review of evidence on young people's views and on interventions in mental health, physical activity and healthy eating*, Volume I, Overview; Volume 2, Complete Report, London, EPPI-Centre.

Smith, A.F.M. (1996) Mad cows and ecstasy: chance and choice in an evidence-based society, *Journal of the Royal Statistical Society*, 159, 3, 367–383.

Stanley, J.C. (1957) Controlled experimentation in the classroom, *Journal of Experimental Education*, 25, 195–201.

Stewart, R. & the HIVSA Team (eds) (2001) *The HIVSA Project Report*, London, Social Science Research Unit.

Tooley, J. & Darby, D. (1998) *Educational Research—a Critique*, London, Office for Standards in Education.

Torgerson, C., Roberts, B., Thomas, J., Dyson, A. & Elbourne, D. (in press) Developing protocols for systematic reviews in education: early experiences from EPPI-Centre Review Groups, *Evaluation and Research in Education*.