BEME Guide No. 1: Best Evidence Medical Education

R. M. HARDEN[1], JANET GRANT[2] GRAHAM BUCKLEY[3] & I. R. HART[4]

[1]Education Development Unit, Scottish Council for Postgraduate Medical & Dental Education, Dundee, Scotland, UK [2]Open University Joint Centre for Education in Medicine, London, UK [3]Scottish Council for Postgraduate Medical & Dental Education, Edinburgh, Scotland, UK [4]Ottawa, Canada

SUMMARY *There is a need to move from opinion-based education to evidence-based education. Best evidence medical education (BEME) is the implementation, by teachers in their practice, of methods and approaches to education based on the best evidence available. It involves a professional judgement by the teacher about his/her teaching taking into account a number of factors—the QUESTS dimensions. The **Q**uality of the research evidence available—how reliable is the evidence? the **U**tility of the evidence—can the methods be transferred and adopted without modification, the **E**xtent of the evidence, the **S**trength of the evidence, the **T**arget or outcomes measured—how valid is the evidence? and the **S**etting or context—how relevant is the evidence? The evidence available can be graded on each of the six dimensions. In the ideal situation the evidence is high on all six dimensions, but this is rarely found. Usually the evidence may be good in some respects, but poor in others. The teacher has to balance the different dimensions and come to a decision on a course of action based on his or her professional judgement. The QUESTS dimensions highlight a number of tensions with regard to the evidence in medical education: quality vs. relevance; quality vs. validity; and utility vs. the setting or context. The different dimensions reflect the nature of research and innovation. **B**est **E**vidence **M**edical **E**ducation encourages a culture or ethos in which decision making takes place in this context.*

The need for evidence-based teaching

There can be few subjects, if any, where there is as great a degree of internal dissension as education (Squires, 1999). There are tensions as to what is taught and how it is taught, with the curriculum destined, many would argue, to remain an area of conflict. In medical education, change is very much on the political, professional and public agenda. Reports from bodies such as the General Medical Council (1993) in the UK, the World Federation for Medical Education (Walton, 1993), and the Association of American Medical Colleges (1994, 1998; Anderson & Swanson, 1993) in the USA argue powerfully for revisions to the medical curriculum and for changes in teaching practices. Individual teachers engaged in undergraduate, postgraduate and continuing education are caught up and struggle with this movement for change. Will a new approach that has been advocated work in their practice and will it prove to be better or worse than what they are currently doing? Does the adage that new is better apply in their case? "It is often unclear", Davies (1999) concluded, "whether developments in educational thinking and practice are better, or worse, than the regimes they replace". New approaches may be introduced in medical education with much rhetoric but little real, reliable or valid evidence. Other teachers may follow in lemming-like droves before evidence is available confirming the value of the approach, and find themselves locked in, with evangelical partisanship determining action.

So education often develops and changes simply on the basis of new ideas promoted with missionary zeal, new theories with very little evidential basis and the social and political values of the moment. Very often, ideas which have no evidential basis become so ingrained by constant repetition and reassertion that the emperor's new clothes almost seem to be real.

Thus we need to think more critically about current educational practice and about new approaches to medical education. The need for evidence-based medical education is highlighted in editorials in *Medical Teacher* (Harden, 1998; Hart, 1999), and in the *British Medical Journal* (Petersen, 1999), which suggests that "the evidence base is as important in educating new doctors as it is in assessing a new chemotherapy", "Ultimately research into teaching and learning in medicine", argue Bligh & Parsell (1999), "has its impact at the bedside, in the consulting room and in the wider community. Research in medical education matters."

Problems with evidence-based teaching

There is, however, a problem. Van der Vleuten (1995) highlighted a paradox in medical education:

> I noticed that my new colleagues—clinical and biomedical researchers—had the same academic values as I did, which reassured me and made me feel comfortable. However, I quickly noticed something peculiar; the academic attitudes of the researcher appeared to change when educational issues were discussed. Critical appraisal and scientific scrutiny were suddenly replaced by personal experiences and beliefs, and sometimes by traditional values and dogmas.

There is a widely held view among clinicians, medical researchers and medical teachers that evidence to support (or reject) educational approaches is not available. This may be true in some areas but not in others. In the area of teaching and learning communication skills in medicine, Aspegren (1999) identified 180 pertinent papers including

Correspondence: Professor R. M. Harden, Centre for Medical Education, Tay Park House, 484 Perth Road, Dundee DD2 1LR, UK. Tel: +44 (0)1382 631972; Fax: +44 (0)1382 645748; Email: p.a.wilkie@dundee.ac.uk

ISSN 0142-159X (Print) 1466-187X (Online)/99/060553-10

31 randomized studies. Powis (1998) studied approaches to student selection and described an evidence-based Admissions Process at Newcastle (New South Wales) Medical School. "There is a huge body of research evidence out there but it is either not known about or ignored", suggests Gibbs (1995). "It is hard to imagine what further research on lecturing, for example, could make any difference to the business of changing compulsive lecturers' minds." Evidence is, however, frequently ignored (Hargreaves, 1996), and there is, at present, a gap between educational researchers and users of educational research. Often those who are concerned about a lack of evidence either have not looked or have looked in the wrong places. Campbell & Johnson (1999), for example, concluded, on the basis of a literature survey restricted to Medline, that there was no evidence to support multi-professional or multimedia education. Such a restricted literature survey excludes many research studies that address these areas. Lack of evidence should not be used by teachers as an excuse for a failure to adopt an evidence-based approach to their teaching practice.

In medicine, evidence-based practice has been widely accepted and has been defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett *et al.*, 1996). Since its inauguration in 1993, the international Cochrane Collaboration has grown to consist of about 50 Collaborative Review Groups whose members are preparing and maintaining systematic reviews of the effects of healthcare interventions (Chalmers *et al.*, 1997). Why are the same principles not applied to teaching? It has been argued that there are problems of measurement and causation in educational research that are not found in medicine. Labaree (1998) contrasts the hard knowledge of the natural sciences with the soft knowledge produced by the humanities and the social sciences: "Researchers and practitioners in these areas pursue forms of enquiry in which it is much more difficult to establish findings that are reproducible and where validity can be successfully defended against the challenge of others." Compared with medicine, research in education may be more complex, confounding factors may be more apparent, content may be more implicit and controlled trials may be difficult. Moreover the impact of education on patient care and the health of the community is less direct than with medical interventions such as a new drug or surgical procedure (Figure 1). Indeed, Campbell & Johnson (1999) suggest that "The epistemological assumptions underlying evidence-based medicine are inappropriate for medical education. The resulting straitjacket would severely limit the expression of medical education research and practice . . . ."

Many would disagree with this view and Davies (1999) has argued that, when compared with medicine, education faces very similar, if not identical, problems of complexity, context specificity, measurement and causation. Many of the problems about the complexity of education and social interventions and their evaluation apply to health care too. "It is just", suggests Oakley (1999), "that health care is conventionally portrayed as simpler". In medicine, for example, those interested in the management of stroke were "lulled into intellectual complacency by an uncritical acceptance of analogies with myocardial infarction" (Ellis & Matthews, 1999). Meta-analyses demonstrated that organized stroke unit care reduces death, dependency and the need for institutional care. It is not clear, however, what it is about organized care and which patients benefit, as meta-analysis failed to find any sub-group of patients or model of stroke unit care particularly associated with benefit. Another example of the complexity of evidence-based practice in medicine is the notion that sunlight is bad for your health, a view that has been widely embraced by doctors mainly on the basis that exposure to the sun increases the incidence of malignant melanoma. This ignores, however, that increased exposure to sunlight may have beneficial effects in other diseases (Ness *et al.*, 1999). One should not simply dismiss, therefore, the idea of evidence-based teaching on the grounds that it is more complex than evidence-based medicine.

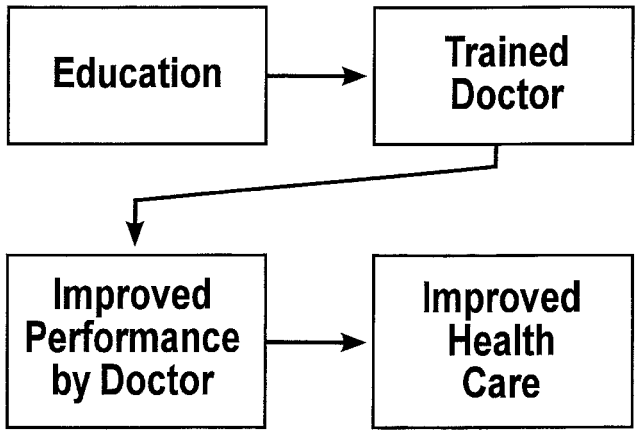

Figure 1. A problem with evaluating the value of an educational approach is that education is at some distance from the ultimate target—improved healthcare in the community.

The concept of best evidence medical education (BEME)

Given these problems, it is not surprising that opinion about the application of the findings of research in medical education is polarized, with the choice presented as 'evidence-based' teaching or 'opinion-based' teaching (Figure 2). A more helpful view of evidence-based teaching is of it as a continuum between 100% opinion-based education at one end of the spectrum where no useful evidence is available, and 100% evidence-based education at the other where decisions can be taken on the basis of detailed evidence (Figure 3). In best evidence medical education (BEME), teachers make decisions about their teaching practice on

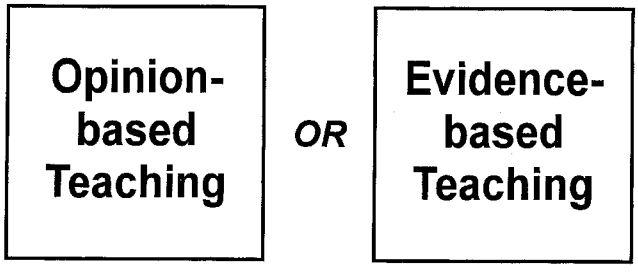

Figure 2. The choice may be presented as opinion-based or evidence-based teaching.

the best evidence that is available at whichever point they find themselves on the continuum. In topics such as teaching and learning about communication skills where a significant body of evidence is available (Aspegren, 1999), the teacher should be nearer the right on the continuum. In other areas such as setting the optimum duration of postgraduate training, the evidence is less clear-cut and we are nearer to the left on the continuum. In best evidence medical education, the culture or ethos is such that teachers are encouraged to question their practice, to look for the best evidence available, to relate the evidence to their own situation and to apply their professional judgement. Hart (1999) has suggested that "Taking a best evidence-based approach to medical education questions forces educators to:

(1) comprehensively critically appraise the literature that already exists in the area, and categorise the power of the evidence available, and
(2) identify the gaps and flaws in the existing literature and suggest (and if possible carry out) appropriately planned studies to optimize the evidence necessary to make the proposed educational intervention truly evidence based."

The teacher can be assisted to identify the evidence available through a study of systematic literature reviews and access to appropriate databases. Given that the quality, relevance and validity of the evidence are likely to be variable, the question arises as to how the teacher can be assisted to evaluate the evidence for relevance to his/her own practice. This is more important than the more elitist and less appropriate question which is sometimes asked: How can research workers influence the behaviour of practising teachers? In best evidence medical education the individual teacher makes his or her decisions on the best evidence available. In some instances the evidence may be more to the left of the continuum, in others more to the right. With increased activity in research in medical education we can expect a movement to the right.

The grading of evidence

There are obvious advantages in a scoring or grading scheme which places the evidence available at the appropriate point on the continuum between opinion-based and evidence-based teaching. In evidence-based medicine, a grading of the evidence used by the US Agency for Health Care Policy and Research and adopted by the Scottish Intercollegiate Guidelines Network (SIGN) (1999) is given in Table 1.

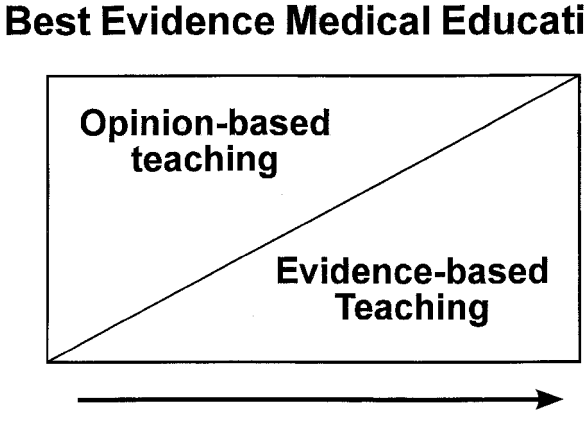

Figure 3. Best evidence medical education can be represented as a continuum between 100% opinion-based and 100% evidence-based education.

Table 1. An example of the definitions of the evidence used by the US Agency for Health Care Policy and Research and the Scottish Intercollegiate Guidelines Network (SIGN).

(1)	Evidence obtained from meta-analysis of randomized controlled trials
(2)	Evidence obtained from at least one randomized controlled trial
(3)	Evidence obtained from at least one well-designed controlled study without randomization
(4)	Evidence obtained from at least one other type of well-designed quasi-experimental study
(5)	Evidence obtained from well-designed non-experimental descriptive studies, such as comparative studies, correlation studies and case studies
(6)	Evidence obtained from expert committee reports or opinions and/or clinical experiences of respected authorities

We have explored the development of a similar grading scheme as a basis for evaluating research in education:

0 No evidence
1 Evidence-based on professional judgement
2 Evidence based on educational principles
3 Evidence based on experience and case studies
4 Evidence based on consensus views built on experience
5 Evidence based on studies in a comparable but not identical area
6 Evidence based on well-designed non-experimental studies
7 Evidence based on well-designed quasi-experimental studies
8 Evidence based on well-designed controlled studies

In practice, such a grading scheme proved difficult to use. It soon became obvious that the continuum was multi- and not unidimensional as was implied by the eight-point grading scheme. The unidimensional approach was replaced by a multidimensional approach with six dimensions, each with its own continuum, and represented by the QUESTS acronym (Table 2).

Table 2. The QUESTS dimensions for evaluating evidence in educational practice.

1. Quality	How good is the evidence?
2. Utility	To what extent can the method be transferred and adopted without modification?
3. Extent	What is the extent of the evidence?
4. Strength	How strong is the evidence?
5. Target	What is the target? What is being measured? How valid is the evidence?
6. Setting	How close does the context or setting approximate? How relevant is the evidence?

Dimension 1—the quality of evidence

What counts as evidence? In research a major emphasis is placed on quality and on controlled experimental studies. This is illustrated in the evidence-based medicine grading given in Table 1. Indeed, it is sometimes suggested that only randomized controlled trials (RCTs) should be included in reviews of research. Randomized controlled trials, however, are difficult to undertake in evaluations of teaching or learning effectiveness, though their potential has been recognized by some researchers (Boruch, 1997). In education research there may also be an over-emphasis on quantitative methods and a failure to recognize the relevance of qualitative methods (Harden, 1986).

Relevant evidence, however, may come not from formal experimental or quasi-experimental research studies but from professional experience and professional judgement. In education these may be important sources of evidence. Research data may not be available in many areas of education, but approaches to education may have been tried and tested in the natural experience of medical education over the years.

Theory or educational principles may also inform the development and evaluation of educational interventions. "Interventions", suggests Oakley (1999), "may be based on prior evidence about what works, on guesswork, individual practitioner preferences, and/or the usual *a priori* enthusiasm for innovation; but some interventions, especially in the social field, are informed by theories about processes of intervening and/or bringing about behaviour change". The extent to which a theory makes a difference to the effectiveness of an intervention, however, remains to be evaluated.

Points on a quality continuum might include:

(1) evidence based on professional judgement—the beliefs and values of experienced teachers;
(2) evidence based on educational principles;
(3) evidence based on professional experience;
(4) evidence based on case studies;
(5) evidence based on cohort studies and related methods;
(6) evidence based on randomized controlled trials.

There are dangers, however, in thinking about the quality of evidence in terms of a hierarchy of methods as suggested in this list. Other factors may adversely affect the quality or robustness of the study. It needs to be recognized that each approach to educational research has its own advantages, indications and, most importantly, limitations. There are, for example, no randomized controlled trials which prove the link between smoking and cancer, nor are there likely to be. The results of a large scale RCT, if available, may be helpful at the point of deciding whether to adopt an educational approach or not. Evidence about how to implement the approach, however, may be better obtained from well-documented case studies. Marian Warnock (1994), in the Gifford lectures given in Glasgow in 1992, drew attention to the role of education in transferring values from one generation to another, aspects of education which are intrinsically more difficult to measure than the more technical competences.

Second, it is important to recognize that the method by itself does not guarantee the quality of a study. Questions which should be asked of research or evaluation evidence are given in Table 3. These can be used as a basis for assessing the quality of a research-based piece of evidence.

Where there is little or no research-based evidence, we have to use our independent and professional judgement to decide whether the idea is a good one or not for medical education. It is necessary and wise, however, to ask questions of the ideas that are put forward and even put into use on the basis of no known evidence. Sometimes you will conclude that they are good and sometimes you will conclude the opposite. The questions noted in Table 4 may help you to assess the quality of evidence based on experience, opinion or theory.

Dimension 2—the utility of the evidence

The utility of the evidence is the extent to which the method or intervention, as reported in the original research report, can be transplanted to another situation without adaptation. Rank Xerox, a leader in knowledge sharing, admits it lost numerous best practices because people tweaked them before implementing them (Rank Xerox, 1998). Antil *et al.* (1998) looked at the widespread adoption of cooperative learning in schools. They reported that "the majority of teachers were using a form of cooperative learning that differed from those described by researcher-developers". It is difficult to be certain in these circumstances whether the benefits of cooperative learning found by the original researchers will be transferred to the teachers' practice.

Will changes in the number of students in a PBL group affect the conclusions about the effectiveness of the method? An increase from 6 to 7 or 8 may not. An increase to 10 or 12 is more likely to do so. Will conclusions about the value of computer-assisted learning be affected if the circumstances of the original study in which each student was required to have their own laptop computer do not apply? Will conclusions about the role of interviews in the selection of students for admission to a medical course be affected by minor changes in the composition or training of members of the interviewing committee?

One cannot always predict the effect of changes made to the method as originally reported. In Dundee, a self-learning laboratory in biochemistry in which students used audiovisual learning programmes and other resources was successfully introduced and was popular with students and staff (Macqueen *et al.*, 1976). A feature of the learning area was the presence of a student-friendly member of staff whose responsibility it was to facilitate the students' learning. The model of self-learning was copied in a number of other institutions, but often with modifications. In one school, where the approach had been adopted, the method was found to be less popular with students. This may have been the result of substituting a computer management system for the staff facilitator.

Changes to procedures or to a method may have positive rather than negative effects. Many reports have documented problems related to lectures and the students' passive role in the learning process. The situation can be changed dramatically by incorporating student participation in the lecture. The lecturer may, for example, address a question to the class. Two or three students sitting adjacent to each other are then required to discuss the question and agree an answer, which they signal using a remote feedback device. This changes the character of the lecture and its educational

Table 3. Questions to ask of research or evaluation evidence.

Area	Questions	Yes	No	N/A
Background	Is the research free of theoretical views already held by the authors?			
	If the evidence is based on cited papers, are those papers researched based rather than theory only?			
	Are the researchers independent?			
Sample	Is it large enough for the purpose?			
	Could we safely say something about the general case on the basis of this sample?			
	Is there a reasonable response rate?			
	Is the sample biased in any way?			
Data collection	Do you know how the data were collected?			
	Is the data collection instrument properly described?			
	Was the data collection instrument properly developed and piloted or tested?			
Data analysis	Is the way the data were analysed properly described *so that you could do it in the same way*?			
Validity, reliability and generalizability	Did the study try to establish the validity of the data and findings?			
	Did the study try to establish the reliability of the data and findings?			
	Is the likely generalizability of the study discussed?			
Conclusions	Are the conclusions reached actually borne out by the data?			
	Do the recommendations actually follow on from the findings?			
	Does the research justify the conclusions? Eg small numbers in a qualitative study should not merit general conclusions for action.			

Table 4. Questions to ask of evidence based on experience, opinion or theory.

Questions to ask:	Yes	No	N/A
Would the approach be accepted by most informed/respected practitioners in the field?			
Is the view put forward by a practitioner who understands the field?			
Does the view seem to take account of what is special to medical education?			
Is the view based more on the practice than theory?			
Is the view derived from medicine or a closely related field?			
If the view propounds a theory, was that theory developed in medical education?			
Is there a commonly recognized good reason for adopting the view?			
Does the view seem to be more than rhetoric, i.e. more than an often repeated statement that is now the received wisdom?			
Is the view based on political or social values that are of central importance to medicine?			
Is the view a practical one based on the context of medical education?			
Does the view make sense, i.e. in your professional judgement, does it seem to have face validity?			
Does the view seem to fit on with the professional values of medicine?			

potential. With this alteration the 'lecture' can become a powerful educational tool.

It may be expected that as experience is gained with an educational approach and modifications made to it, the approach will be more effective. Grant (1999) has suggested that health technologies that change during evaluation are a challenge to health technology assessment. There is a reluctance to evaluate these technologies until they are used in a standardized way. A particular component of technology change is 'learning' such as seen during the adoption of keyhole laparoscopic surgical techniques. The same learning may be a feature of new approaches in education. There may be the expectation, but not the certainty, of improved results with changes made to the original educational approach described.

Dimension 3—extent of the evidence

What is the extent of the evidence available? Is the evidence based on a single opinion or study of an isolated example of

the approach working well, on a consensus view, a systematic review of the literature, or on a meta-analysis of a number of studies? The danger of educational research relying on the results of single studies has been emphasized (Foster & Hammersley, 1998). Meta-analysis is essentially a form of literature review that summarizes the features and outcomes of a body of research or in statistical terms is "a statistical synthesis of the numerical results of several trials that all examined the same question" (Greenhalgh, 1997). One has to look critically, however, at meta-analysis to ensure that like is being compared with like and spurious conclusions are not being drawn from the data available. "Meta-analysis in educational research has the same problems as in health care research, such as ensuring the comparability of different samples, research designs, outcome and process measures, identifying confounding factors and bias, and determining the attributable effects of the intervention(s) being assessed" (Davies, 1999).

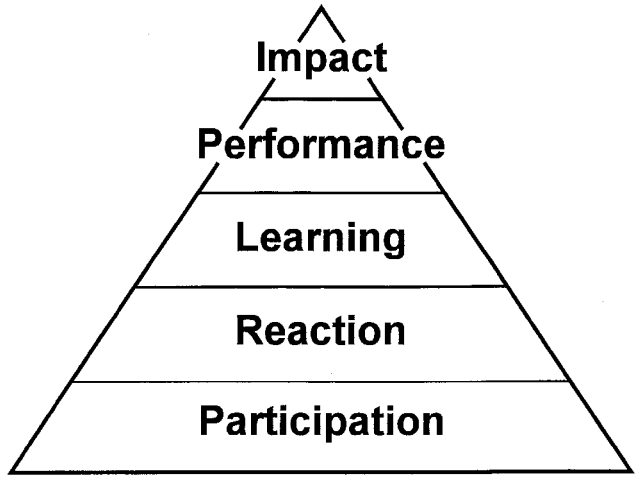

Figure 4. A modified version of Kirkpatrick's hierarchy of levels of evaluation.

Dimension 4—the strength of the evidence

In some instances the effect of an intervention may be obvious. Studies of the use of simulated patients, for example, have clearly demonstrated that subjects could be trained to act as patients in clinical examinations and could not be distinguished by students or examiners from 'real' patients (Collins & Harden, 1998). Sometimes, however, conclusions may be drawn on the basis of less strong evidence. The effect of the provision of written information on patient satisfaction in student examinations was studied by Welfare *et al.* (1999). They recommended that all patients attending for medical examinations should be provided with written information. While many workers in the field would concur with this conclusion, the evidence presented was not strong, with a p value of 0.077. In evaluating evidence, one needs to be more critical of evidence in which the results have only marginal statistical significance.

Dimension 5—the target for the evidence

The validity of a research study depends on the questions it addresses. The inappropriateness of the answers from a study may mean simply that we have asked the wrong questions. Critical to any evaluation of the relevance of a research study is the nature of the outcome or the target that was assessed. A large sample size may increase the probability of statistically significant findings even though the practical significance of these findings may actually be negligible. Guglielmi & Tatrow (1998) described examples of conceptually trivial but significant correlations in the field of research into teacher stress and burnout.

Kirkpatrick (1967) has described a hierarchy of levels of evaluation and a modified version is show in Figure 4. At the bottom of the pyramid are studies that look only at participation in an education activity. How many doctors attended the continuing education programme? How many students used the computer-based learning programme? How much reading on the topic did students complete? Other studies have looked at the learners' reactions to the programme. Did they feel they learned from it? Was it easy or enjoyable to use? Did they wish further learning programmes presented in the same way? Such information may be obtained using techniques such as questionnaires, structured interviews or focus groups. Other studies have looked at the learning gains. What new knowledge have the students gained? What practical skills have they acquired? Have their attitudes changed as a result of the intervention? It may be more important to identify whether any changes in the learners' performance or behaviour have resulted from the intervention.

At the top of the hierarchy are studies that look at the impact of education on the delivery of healthcare. While this is a highly desirable target to aim for, it is difficult to assess. Is an educational package on the treatment of hypertension reflected in the doctors' management of patients with a raised blood pressure? Do doctors who complete an educational programme make fewer errors in practice in the area covered by the programme than a control group who have not? Does an educational programme result in patients who are more satisfied with their management? There are two main considerations in relating education intervention to the assessment of outcomes of care, according to Tamblyn (1999): "The first consideration is the population impact of optimal, average or suboptimal medical practice. The second consideration is the strength of the inference that can be drawn about an individual physician's contribution to the standard of care received and/or the resulting health outcome."

There is not always a good correlation between these different outcomes measured. Educational outcomes, as measured by performance in examinations, may correlate only poorly with educational outcomes as reflected in changes in practice (Gonella *et al.*, 1994, Rethans & van Boven (1987). Oswald (1999) has highlighted that, fundamentally, medical education should be concerned with improving patient care. He suggests that: "When we are able to firmly connect innovation and quality in education with better outcomes for patients, then we shall be taken seriously in RAE [Research Assessment Exercise] terms." As one moves up the hierarchy of outcomes of an evaluation, however, the situation becomes more complex. There are more confounding factors and evaluation is more time consuming.

Another problem with education is that different educational goals may be emphasized for the same intervention (Donmoyer, 1985). This may result in conflicting criteria

for evaluating the educational research. For example, a concentration on knowledge gain may militate against a change of behaviour in practice.

In evaluating evidence, perhaps the most important factor is the target of the research or the outcomes measured. The validity of the evidence and what is being measured is of the greatest importance. It is the quality of the benefit that matters, perhaps even more so than the quality of the research and the size or even the certainty of the likelihood.

Dimension 6—the setting of context of the evidence

Teaching and learning takes place in a range of settings or contexts:

- different phases of education—primary school, secondary school, higher education, postgraduate education and continuing education. How applicable, for example, are the results of an evaluation of computer-assisted learning in a first-year medical programme to the continuing education of consultant physicians?
- different professions or disciplines within the same profession. Can one assume that an approach to problem-based learning effective with medical students will also be effective with nursing students or vice versa?
- different ages and sex distributions in the subjects studied. Gender differences are well recognized as a potential confounding factor in educational research.
- different geographical or cultural backgrounds. Can approaches to the use of lectures or to ethical training be transferred from one culture to another?

Reed & Proctor (1995) describe how research deals with ambiguity and messy context-dependent problems. Research may show that a method or approach works. This may apply, however, only in a particular context or set of experimental conditions. Whether one medium for a learning package proves more effective than another may be more dependent on the expertise of the instructional designers rather than on intrinsic differences between the two media (Harden, 1986). "Can we assume", asks Hammersley (1997), "that causation in this field involves fixed universal relationships, rather than local, content sensitive patterns . . . ." Labaree (1998) suggests that: "The only causal claims educational research can make are constricted by a mass of qualifying clauses, which show that these claims are only valid within the artificial restrictions of a particular experimental setting or the complex peculiarities of a particular natural context."

The setting or context of educational research is important. Teaching practice is context and culturally specific, and research findings in one area may be of limited value to those in different practice settings. There is no such thing as context-free evidence (Davies, 1999).

Because of the importance of setting, some may argue that teaching is a series of different jobs that are dependent on the practice settings. Such differences must be taken into account when evaluating the transferability of the findings of educational research. One need not be too depressed, however, by differences in context or setting. It can be argued that many of the basic principles of education apply almost regardless of the setting.

Use of QUESTS in best evidence medical education

It is very difficult to undertake meaningful research in education. The variables are too diffuse and difficult to identify. Very often they are not easy to measure. Other factors often contaminate the relationship between an educational event and its eventual outcome. Sometimes, particular outcomes are not easy to specify, nor are the timescales in which we might expect to see an effect or to see an effect last. Despite these difficulties, there is in medical education a growing body of evidence relating to teaching methods, approaches to assessment, curriculum planning and student selection.

The problem is not so much that teachers do not undertake research (although more research is needed), but that there is not a culture of teachers using research to inform their teaching practice (Davies, 1999). The aim of best evidence medical education is to have medical teachers think more clearly about the actions they are taking as teachers and to utilize evidence where it is relevant and available to inform their decision. The practice of medical education is currently a scene of great activity. This is not always matched, however, by an understanding of the actions, and even less frequently is evidence relating to the action considered (Figure 5).

In best evidence medical education teachers combine their teaching and professional judgement with the evidence available in order to decide the most appropriate action in a particular situation. The QUESTS continuum can assist with this process. In the ideal situation, the evidence available would be to the right on all dimensions as indicated by C in Figure 6. The evidence would be of high quality and it would have a high utility. The results from multiple studies would be available and the evidence that existed would be strong and not weak. The evaluation of the intervention would include an assessment of changes in healthcare delivery that resulted from the educational intervention. Finally, the setting or context in which the evidence was collected would approximate to that of the teacher. In contrast, a reference quoted in support of a particular stance may be to the left on all dimensions as shown in A in Figure 6 with low-quality, low-utility evidence which is based on a single study and is lacking in strength. The target or outcome

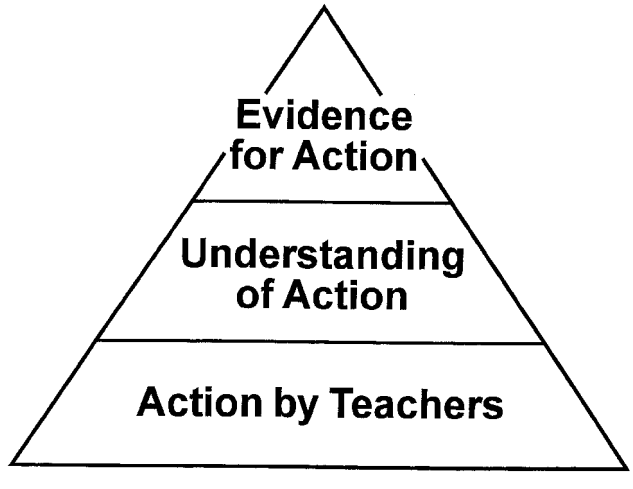

Figure 5. In the practice of medical education the actions taken by the teacher are often not matched by an understanding of the actions and even less frequently by research evidence in support of them.

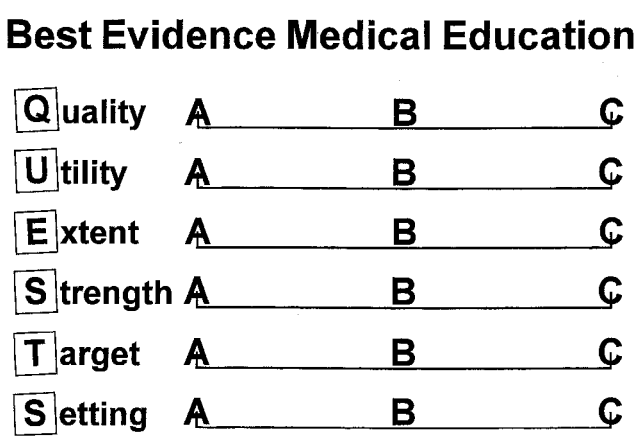

Figure 6. Research studies may be placed to the right of the continuum as at C, to the left as at A, or somewhere in the middle as at B.

evaluations are at the bottom of the hierarchy with participation used as a measure of success. The evidence was obtained in a different context.

It is more likely, however, that the study will be somewhere between the two extremes as shown at B. It is even more likely that the position between the two extremes will vary with the six different dimensions as shown in Figure 7.

Some of the dimensions are intrinsic to the source of evidence. This applies to the quality of the study and the methodology used, the extent and strength of the study and the outcomes measured. Other dimensions such as the utility of the study and the context or setting are a function of the extent to which the teacher can relate the study to his or her own context.

Professional judgement by the teacher is needed to draw conclusions about the evidence as described in the six QUESTS dimensions. This is not a problem where the point on the continuum is similar in all six settings, as in Figure 6. It is more difficult as in Figure 7, where the point varies from one continuum to another. In this situation the teacher's judgement is needed to integrate and balance the different scales.

Tensions highlighted

A number of tensions in applying evidence-based teaching are highlighted by the QUESTS dimensions. There are tensions between the quality of the evidence and its applicability in a particular setting. "There is the lack of acknowledgement", suggest Perkins *et al.* (1999), "of the uniqueness of practice contexts . . . . Where research does say unequivocally that X works, and it frequently does not, this often applies only in a set of experimental conditions which are not reproduced in most real life settings." Should more or less importance be attached, for example, to a single randomized controlled trial carried out in a different setting compared with a series of carefully reported case studies carried out in a similar setting? A similar problem has been identified with the evaluation of research in patient education by Herbert (1998):

> Researchers and clinicians struggle with the application of research findings to their own patients and settings. Often, results that are reported are short term in sample populations that are highly selected. The questions we ask ourselves are 'can I apply this in my community, in my practice, in my institution and get similar results?' and 'if this method works for condition X, will it also work for condition Y?' Often, the answer to both questions is, we really do not know, as the research has not been done. Our only recourse is to try to choose approaches that are based on sound theoretical models, to reflect critically on the outcomes in our particular situations, and to modify the method to 'fit' the local situation.

The teacher's knowledge and understanding of their context needs to be considered alongside evidence from research studies carried out in other contexts. "Those who ignore it (practitioner knowledge) in pursuit of evidence-based practice based purely on evidence collected through scientific or social scientific methods", suggest Perkins *et al.* (1999), "will probably find that their schemes fail."

There is also a tension between the push for higher-quality research and controlled trials, often at the expense of validity and the targets or outcomes evaluated. There is a risk that pressures for more quality may promote a narrow perspective of educational research where there is more high-quality research but more trivial or less relevant conclusions. There is a risk in the search for a rigorous, robust quality evaluation that one ignores the crucial point of what is being evaluated.

A further tension exists between the utility and the setting dimensions. The teacher tends to compensate for a difference between the setting in which the research was undertaken and the context in which he or she practises by adapting the method to suit the local context. This inevitability, however, lowers the utility scale.

Conclusion

The adoption of best evidence medical education does not require the teacher to be a researcher in education. It does require the teacher to be able to appraise the evidence available and come to a decision on the basis of his or her clinical judgement. The process may also highlight areas where there is a need for further research. Best evidence medical education is an attitude of mind. It involves the creation of a culture or ethos in which teachers think critically about what they are doing, look at the best evidence

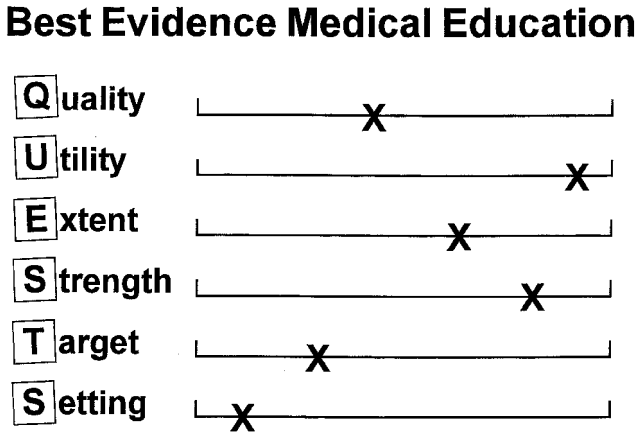

Figure 7. The evidence available may lie at different points between the extremes on each dimension.

available and on the basis of this make decisions about their teaching practice. The teacher in a traditional school may question the role of teaching methods such as the lecture and the teacher in a PBL school may question the role of the teacher in the group process adopted.

How much evidence is required before the teacher should act cannot be stated with any certainty. In particular because of the context dependent nature of education, the evaluation of an approach in a particular context must depend heavily on the experience of the teacher in that context. What is good enough evidence will depend on:

- the cost of the implementation;
- the problems associated with difficulties that may arise;
- the flexibility of the innovation and the extent to which mistakes can be corrected subsequently.

QUESTS offers a model which helps the teacher or the institution to make decisions about their teaching, taking into account a range of relevant factors in the context of their own teaching practice. Best evidence medical education occurs when decisions relating to teaching are taken with due weight accorded to all valid relevant information on the QUESTS dimensions. Best evidence medical education creates an opportunity for improved teaching by engaging the teacher in the decision process, not by providing him or her with a cookbook of recipes. The approach described also has immediate relevance to the planner or educational administrator, and provides them with a powerful tool to move forward the best evidence medical education agenda.

Brown (1996) has outlined her view of what research can and cannot do:

> It can help our understanding about how things are . . . and why they are the way they are. It can articulate the ways in which they might be different and alternative actions or decisions which might be taken to achieve change. It can elaborate on the implications of making choices among those alternatives. What it cannot do is tell policy makers or practitioners what they should decide or what they should do.

Best evidence medical education places the decision making by the teacher in the context of the best evidence available at the time.

In medicine and in other academic areas, there is some concern that staff activities in teaching are regarded as in some way inferior to research activities. There is a recognized need to improve the image of teaching and to value more highly the wide range of activities in which a teacher is engaged. Active engagement by teachers in the *use* of research through the practice of best evidence medical education may help to address this problem (Sebba, 1999). Best evidence medical education has much to offer the teacher, the student, the medical profession and the public.

Notes on contributors

R.M. Harden is Director of the Centre for Medical Education and Teaching Dean in the Faculty of Medicine, Dentistry and Nursing at the University of Dundee. He is also Director of the Education Development Unit (Scottish Council for Postgraduate Medical & Dental Education), Dundee, UK.

Janet Grant is Professor of Education in Medicine at the Open University and The Joint Centre for Education in Medicine. She is an international expert on medical education with a broad range of research and publications in the field.

Graham Buckley is Director of The Scottish Council for Postgraduate Medical and Dental Education, Edinburgh, UK.

I.R. Hart is Professor Emeritus of Medicine at the University of Ottawa, Canada, with over 25 years' experience of international medical education.

References

Association of American Medical Colleges (1994) Physicians for the Twenty-first Century: Report of the project panel on the General Professional Education of the Physicians and College Preparation for Medicine, *Journal of Medical Education*, 59, part 2.

Association of American Medical Colleges (1998) *Report 1: Learning Objectives for Medical Student Education. Guidelines for Medical Schools* (Washington, DC, Medical School Objectives project, AAMC).

Anderson, M.D. & Swanson, A.G. (1993) Educating medical students—the ACME-TRI report with supplements, *Academic Medicine*, 18, pp. 284–297.

Antil, L.R., Jenkins, J.R. & Wayre, S.K. (1998) Co-operative learning research: prevalence, conceptualisations and the relation between research and practice, *American Educational Research Journal*, 35(3), pp. 419–454.

Aspegren, K. (1999) Teaching and learning communication skills in medicine: a review with quality grading of articles, *Medical Teacher*, 21(6), in press.

Bligh, J. & Parsell (1999) Research in medical education: finding its place, *Medical Education*, 33, pp. 162–164.

Boruch, R.F. (1997) *Randomised Experiments for Planning and Evaluation: A Practical Guide* (Thousands Oaks, CA, Sage Publications).

Brown, S. (1996) *The Role of Research in Mature Education Systems: Proceedings of the NFER International Jubilee Conference*, Oakley Court, Windsor.

Campbell, J.K. & Johnson, C. (1999) Trend spotting: fashions in medical education, *British Medical Journal*, 318, pp. 1272–1275.

Chalmers, I., Sackett, D. & Silagy, C. (1997) The Cochrane collaboration, in: A. Maynard & I. Chalmers (Eds) *Non-random Reflections on Health Services Research*, pp. 231–249 (London, British Medical Journal Publishing).

Collins, J.P. & Harden, R.M. (1998) AMEE Medical Education Guide No 13: real patients, simulated patients and simulators in clinical examinations, *Medical Teacher*, 20(6), pp. 508–521.

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

Donmoyer, R. (1985) The rescue from relativism: two failed attempts and an alternative strategy, *Educational Researcher* 14(10), pp. 13–20.

Ellis, S.J. & Matthews, C. (1999) What has gone wrong in stroke research?, *Postgraduate Medical Journal*, 75, pp. 449–450.

Foster, P. & Hammersley, M. (1998) A review of reviews: structure and function in reviews of educational research, *British Educational Research Journal*, 24(5), pp. 609–628.

General Medical Council (1993) *Tomorrow's Doctors: Recommendations on Undergraduate Medial Education* (London, General Medical Council).

Gibbs, G. (1995) Research into student learning, in: B. Smith & S. Brown (Eds) *Research, Teaching and Learning in Higher Education*, pp. 19–29 (London, Kogan Page).

Gonella, J.S., Hojat, M., Erdmann, J.B. & Veloski, J.J. (1994) The ultimate argument in medical education: the health outcome, *Changing—Medical Education and Medical Practice*, WHO/EDH/NL, 94(2), pp. 8–9.

Grant, A. (1999) Learning curves and health technologies, *The Sorcerer's Apprentice*, Summer, p. 1.

Greenhalgh, T. (1997) *How to Read a Paper: The Basis of Evidence-based Medicine*, p. 119 (London, BMJ Publishing Group).

GUGLIELMI, R.S. & TATROW, K. (1998) Occupational stress, burnout and health in teachers: a methodological and theoretical analysis, *Review of Educational Research*, 68(1), pp. 61–99.

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

HARDEN, R.M. (1986) ASME Medical Education Research Booklet No 2: Approaches to research in medical education, *Medical Education*, 20, pp. 522–531.

HARDEN, R.M. (1998) Medical Teacher, *Medical Teacher,* 20, pp. 501–502.

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

HART, I.R. (1999) Editorial: Best evidence medical education, *Medical Teacher,* 21(5), pp. 453–454.

HERBERT, C.P. (1998) Creative approaches to patient education, *Patient Education and Counselling*, 35, pp. 81–82.

KIRKPATRICK, D.I. (1967) Evaluation of training, in: R. Craig & I. Mittel (Eds) *Training and Development Handbook* (New York, McGraw Hill).

LABAREE, D.F. (1998) Educational researchers: living with a lesser form of knowledge, *Educational Researcher,* 27(8), pp. 4–12.

MACQUEEN, D., CHIGNELL, D.A., DUTTON, G.J. & GARLAND, P.B. (1976) Biochemistry for medical students: a flexible student-oriented approach, *Medical Education*, 10, pp. 418–437.

NESS, A.R., FRANKELL, S.J., GUNNELL, D.J. & SMITH, G.D. (1999) Are we really dying for a tan?, *British Medical Journal*, 319, pp. 114–116.

OAKLEY, A. (1999) An infrastructure for assessing social and educational interventions: the same or different?, Background paper for meeting at the School of Public Policy, University College London, 15–16 July.

OSWALD, N. (1999) Research in medical education, *Medical Education*, 33, p. 470.

PERKINS, E.T., SIMNETT, I. & WRIGHT, L. (1999) Creative tensions in evidence-based practice, in: *Evidence-based Health Promotion*, pp. 7–11 (New York, Wiley).

PETERSEN, S. (1999) Time for evidence based medical education: tomorrow's doctors need informed educators not amateur tutors, *British Medical Journal*, 318, pp. 1223–1224.

POWIS, I. (1998) An evidence-based admissions process at Newcastle (New South Wales) Medical School, *Education for Health*, 11(3), pp. 409–412.

RANK XEROX (1998) *Bulletpoint*, November, p. 1.

REED, J. & PROCTOR, S. (Eds) (1995) *Practitioner Research in Health Care: The Inside Story* (London, Chapman & Hall).

RETHANS, J.J.E. & VAN BOVEN, C.P.A. (1987) Simulated patients in general practice: a different look at the consultation, *British Medical Journal*, 294, pp. 809–812.

SACKETT, D.L., ROSENBERG, W.M.C., GRAY, J.A.M., HAYNES, R.B. & RICHARDSON, W.S. (1996) Evidence-based medicine: what it is and what it isn't, *British Medical Journal*, 312, pp. 71–72.

SCOTTISH INTERCOLLEGIATE GUIDELINES NETWORK (SIGN). Edinburgh 1998. www.show.scot.nhs.uk/sign/home.htm

SEBBA, J. (1999) Priority setting in preparing systematic reviews, background paper for meeting at the School of Public Policy, University College London, 15–16 July.

SQUIRES, G. (1999) *Teaching as a Professional Discipline* (London, Falmer Press).

TAMBLYN, R. (1999) Outcomes in medical education: what is the standard and outcome of care delivered by our graduates?, *Advances in Health Sciences Education*, 4, pp. 9–25.

US DEPARTMENT OF HEALTH AND HUMAN SERVICES, AGENCY FOR HEALTH CARE POLICY AND RESEARCH (1993) *Acute Pain Management Operative or Medical Procedures and Trauma* (Rockville, MD, The Agency) [Clinical Practice Guidelines No 1. AHCPR Publication No 92–0023: 107.]

VAN DER VLEUTEN, C. (1995) Evidence-based education, *Advances in Physiology Education*, 14(1), p. S3.

WALTON, H.J. (1993) Proceedings of the World Summit on Medical Education, *Medical Education,* 28(Suppl. 1), pp. 140–149.

WARNOCK, M. (1994) *Imagination and Time* (Oxford, Blackwell).

WELFARE, M.R., PRICE, C.I.M., HAN, S.W. & BARTON, J.R. (1999) Experience of volunteer patients during undergraduate examination, *Medical Education*, 33, pp. 165–169.