

thorough examination of the strengths and weaknesses of the study itself.

(b) Does it discuss the strengths and the weaknesses of the study in relation to other studies?

Strengths and weaknesses should then be discussed in relation to previous studies. Any differences in results, and why different conclusions have been reached, should particularly be emphasised.

(c) Does it discuss the meaning and implications of the results?

Next, the “meaning” of the study in terms of possible mechanisms, and implications for clinicians or policy makers, needs to be explored but should not be overstated.

(d) Does it discuss the need for further work?

Finally, questions that remain unanswered and what

future work is needed should be discussed without being speculative.

Contributors: Luisa Dillner, Linda Hutchinson, Jean Kerr, Sam Leinster, Kwee Matheson, Stewart Peterson, Mark Rake, Tessa Richards, Richard Smith, and Diana Wood were on the advisory group that formulated the new guidelines and they, along with Peter Cantillon, reviewed an early draft, which was prepared by Kamran Abbasi, Linda Hutchinson, Richard Smith, and Diana Wood. The guidelines were edited by Kamran Abbasi and Richard Smith. Richard Smith is guarantor.

- 1 Altman DG. Randomisation. *BMJ* 1991;292:810-2.
- 2 Altman DG. Better reporting of randomised trials: the CONSORT statement. *BMJ* 1996;313:570-1.
- 3 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA* 1996;276:637-9.
- 4 Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999;318:1224-5.

Evaluating and researching the effectiveness of educational interventions

Linda Hutchinson

Members of the medical profession seem reluctant to value research into the effectiveness of educational interventions.¹ One reason for this reluctance may be that there is a fundamental difficulty in addressing the questions that everyone wants answered: what works, in what context, with which groups, and at what cost? Unfortunately, there may not be simple answers to these questions. Defining true effectiveness, separating out the part played by the various components of an educational intervention, and clarifying the real cost:benefit ratio are as difficult in educational research as they are in the evaluation of a complex treatment performed on a sample group of people who each have different needs, circumstances, and personalities.

Methodology

Choosing a methodology to use to investigate a research question is no different in educational research than it is in any other type of research. Careful attention must be paid to the aims of the research and the validity of the method and tools selected. Educational research uses two main designs: naturalistic and experimental.^{2 3}

Naturalistic design

Naturalistic designs look at specific or general issues as they occur—for example, what makes practitioners change their practice, how often is feedback given in primary care settings, what processes are occurring over time in an educational course, what are the different experiences and outcomes for participants, and can these differences be explained? Like case reports, population surveys, and other well designed observational methodologies, naturalistic studies have a place in providing generalisable information to a wider audience.⁴

Summary points

Health professionals are often reluctant to value research into the effectiveness of educational interventions

As in clinical research, the need for an evidence base in the practice of medical education is essential

Choosing a methodology to investigate a research question in educational research is no different from choosing one for any other type of research

Rigorously designed research into the effectiveness of education is needed to attract research funding, to provide generalisable results, and to elevate the profile of educational research within the medical profession

Experimental design

In contrast to a naturalistic research design, experimental designs usually involve an educational intervention. The parallels with clinical interventions highlight the three main areas of difficulty in performing experimental research into educational interventions.

Complex nature of education

An educational event, from reading a journal article to completing a degree course, is an intervention; it is a complex intervention, and often several components act synergistically. Interventions that have been shown to be effective in one setting may, quite reasonably, not translate to other settings. Educational events are multi-faceted interactions occurring in a changing world and involving the most complex of subjects. Many factors can influence the effectiveness of educational interventions (fig 1).

Department of Child Health, St George's Hospital Medical School, London SW17 0RE

Linda Hutchinson, senior registrar in paediatrics

l.hutchinson@sghms.ac.uk

BMJ 1999;318:1267-9

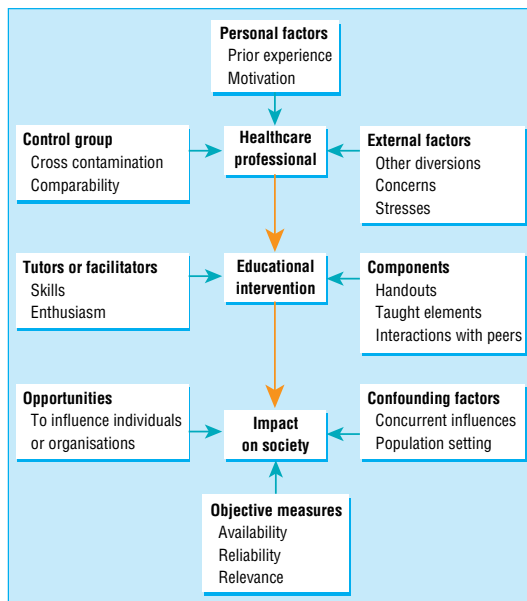


Fig 1 Examples of factors which may influence the effectiveness of educational interventions

Sampling

The randomised controlled trial is regarded as essential to proving the effectiveness of a clinical intervention. In educational research, especially in post-graduate and continuing medical education, the numbers that can be enrolled in a study may not be large enough to allow researchers to achieve statistically significant quantitative results. Comparable control groups may be susceptible to cross contamination from access to some of the elements of the intervention under scrutiny (for example, students may pass their handouts to other students).

Purposive sampling may give different but valid perspectives.³ For instance, if 25 out of 30 people are shown to have benefited from a course, the most interesting question might be “why didn’t the other five people benefit?” Were they demographically different? Did they have different learning styles? More detailed interviews with those five students may be more informative about the philosophy and utility of a course than the simple statistics. Analysing the deviant cases as a project proceeds is important for the validity of results.⁵

Outcome measures

Kirkpatrick described four levels of evaluation in which the complexity of the behavioural change increases as evaluation strategies ascend to each higher level (fig 2).⁶ The length of time needed for the evaluation, the lack of reliable objective measures, and the number of potential confounding factors all increase with the complexity of the change. Researchers in medical education are aware that the availability of funds for research and development is limited unless a link can be made between the proposed intervention and its impact on patient care, yet this is the most difficult link to make.

When does evaluation become research?

Evaluations of the effectiveness of educational interventions may not reach the rigour required for

research even though they may use similar study designs, methods of data collection, and analytical techniques. The evaluation of an educational event may have many purposes; each evaluation should be designed for the specific purpose for which it is required and for the stakeholders involved.^{7, 8} For a short educational course, for example, the purpose might be to assist organisers in planning improvements for the next time it is held. The systematic collection of participants’ opinions using a specifically designed questionnaire may be appropriate. Just as a well designed audit, although not true research, may be informative and useful to a wider audience, so an evaluation may be useful if it is disseminated through publication. But that does not mean that it is research.

For an evaluation of an educational intervention to be considered as research, rigorous standards of reliability and validity must be applied regardless of whether qualitative or quantitative methodologies are used.^{9, 10} For example, if the questionnaires used are not standardised and have not been validated for use in a population similar to the one being studied, they should be piloted and should include checks on their internal consistency. A high response rate—of 80-100%—is required.¹¹ Numerical scores will be meaningless if they are derived from poorly designed scoring systems. If a control group is used, randomisation or case control procedures should meet accepted standards.¹¹

The Cochrane Collaboration module on effective professional practice details quality assessment criteria for randomised studies, interrupted time series, and controlled before and after studies that have been designed to evaluate interventions aimed at improving professional practice and the delivery of effective health care.¹¹ Educational strategies and events for healthcare professionals fall into this remit.

Conclusion

As in clinical research, the need for an evidence base in the practice of medical education is essential for the targeting of limited resources and for informing development strategies. The complexity of the subject matter and the limited availability of reliable, meaningful outcome measures are challenges that must be faced.



Fig 2 Kirkpatrick’s hierarchy of levels of evaluation. Complexity of behavioural change increases as evaluation of intervention ascends the hierarchy

These difficulties are not enough to excuse complacency. Rigorous research design and application are needed to attract research funding, to provide valid generalisable results, and to elevate the profile of educational research within the medical profession.

Competing interests: None declared.

- 1 Buckley G. Partial truths: research papers in medical education. *Med Educ* 1998;32:1-2.
- 2 Patton MQ. *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage, 1987.
- 3 Robson C. *Real world research*. Oxford: Blackwell, 1993.

- 4 Vandenbroucke JP. Case reports in an evidence-based world. *J R Soc Med* 1999;92:159-63.
- 5 Green J. Grounded theory and the constant comparative method. *BMJ* 1998;316:1064-5.
- 6 Kirkpatrick DL. Evaluation of training. In: Craig R, Bittel I, eds. *Training and development handbook*. New York: McGraw-Hill, 1967.
- 7 Edwards J. *Evaluation in adult and further education: a practical handbook for teachers and organisers*. Liverpool: Workers' Educational Association, 1991.
- 8 Calder J. *Programme evaluation and quality*. London: Kogan Page, 1994.
- 9 Kirk J, Miller ML. *Reliability and validity in qualitative research*. Newbury Park, CA: Sage, 1986.
- 10 Green J, Britten N. Qualitative research and evidence based medicine. *BMJ* 1998;316:1230-2.
- 11 Effective professional practice module. In: Cochrane Collaboration. *Cochrane Library*. Issue 2. Oxford: Update Software, 1998.

Evaluating educational interventions

M Wilkes, J Bligh

Recent extensive changes have taken place in medical education at all levels in both the United Kingdom and the United States. These changes need to be assessed to measure how well reforms have achieved their intended outcomes. Educational innovation can be complex and extensive, and its measurement and description is made more difficult by the confounding and complicating effects of each later stage in the continuous curriculum. The radical curriculum reform at undergraduate level in the United Kingdom, managed care in the United States, and the increasing use of community sites for learning in both countries may greatly affect how medicine is practised and managed in the next century.¹ We should know more about the educational processes and outcomes that result from the new courses and programmes being developed in medical schools and postgraduate training.

What is educational evaluation?

Educational evaluation is the systematic appraisal of the quality of teaching and learning.² In many ways evaluation drives the development and change of curriculums (figure). At its core, evaluation is about helping medical educators improve education. Evaluation can have a formative role, identifying areas where teaching can be improved, or a summative role, judging the effectiveness of teaching. Although educational evaluation uses methods and tools that are similar to those used in educational research, the results of research are more generalisable and more value is invested in the interpretation of results of evaluation.

Evaluation can also be a hindrance to curricular change. In the United States, for example, enormous weight is placed on the standardised multiple choice type assessment (USMLE) that is taken by all medical students. Although many people believe in the exam, it has been a major barrier to curricular reform. Medical schools feel that any curricular change may sacrifice students' performance in this examination, which in some circles is still seen as the "gold standard." This reliance on conventional educational tools to compare a new innovative curriculum with the traditional curriculum has caused schools such as McMaster a great deal of angst.

Summary points

Evaluation drives both learning and curriculum development and needs to be given serious attention at the earliest stages of change.

Summative evaluation can no longer rely on a single assessment tool but must include measures of skill, knowledge, behaviour, and attitude

New assessment tools do not necessarily duplicate each other but assess and evaluate different components of a doctor's performance

Assessment needs to be part of an ongoing evaluation cycle intended to keep the curriculum fresh, educationally sound, and achieving its intended objectives

Office of the Dean,
UCLA School of
Medicine, Los
Angeles, CA
90095-7035, USA

M Wilkes,
*senior chair, doctoring
curriculum*

Department of
Health Care
Education,
Liverpool
University Medical
School, Liverpool
L69 3GA

J Bligh,
professor

Correspondence to:
Dr Wilkes
mwilkes@ucla.edu

BMJ 1999;318:1269-72

At this point it is worth differentiating between monitoring, evaluation, and assessment. Assessment refers to the quality measures used to determine performance of an individual medical student. Monitoring is the gathering and recording of data about courses, teachers, or students and is regularly carried out at institutional level. Evaluation uses data gathered in the monitoring process to place a value on an activity. According to Edwards, evaluation seeks to "describe and explain experiences of students and teachers and to make judgements and [interpret] their effectiveness."³

Approaches to evaluation

Recommendations intended to evaluate changing medical programmes have been made in the light of the extensive changes going on in medical schools in the United States.⁴ Four general approaches to educational evaluation have emerged over recent years. We have classified these as follows:

Student oriented—Predominantly uses measurements of student performance (usually test results) as the principal indicator.