THE UNIVERSITY OF
WARWICK

Evaluation – the educational context

C F Macdougall

Updated information and services can be found at:

http://ep.bmj.com/content/95/1/28.full.html

These include:

References	This article cites 7 articles, 5 of which can be accessed free at: **http://ep.bmj.com/content/95/1/28.full.html#ref-list-1**
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To order reprints of this article go to:

http://ep.bmj.com/cgi/reprintform

To subscribe to *Archives of disease in childhood - Education & practice edition* go to:

http://ep.bmj.com/subscriptions

Evaluation – the educational context

C F Macdougall

Correspondence to
Dr C F Macdougall, Medical
Teaching Centre, Warwick
Medical School and University
Hospital, Coventry CV4 7AL, UK;
colin.macdougall@warwick.
ac.uk

ABSTRACT

Evaluation comes in many shapes and sizes. It can be as simple and as grounded in day to day work as a clinical teacher reflecting on a lost teaching opportunity and wondering how to do it better next time or as complex, top down and politically charged as a major government led evaluation of use of teaching funds with the subtext of re-allocating them. Despite these multiple spectra of scale, perceived ownership, financial and political implications, the underlying principles of evaluation are remarkably consistent. To evaluate well, it needs to be clear who is evaluating what and why. From this will come notions of how it needs to be done to ensure the evaluation is meaningful and useful. This paper seeks to illustrate what evaluation is, why it matters, where to start if you want to do it and how to deal with evaluation that is external and imposed.

INTRODUCTION

At the core of good evaluation is a search to show what is good and to improve things. Undertaking evaluation can seem hard work and external evaluation potentially threatening, but if it is good evaluation it will highlight what works well and identify what can be improved. Most clinicians and teachers do what they do because they want to achieve things (make patients better, help a trainee progress, help a department understand a new protocol, etc). Good evaluation ensures that we target our efforts to the maximum advantage and achieve what we set out to achieve, and should therefore be integral to most of what we do.

DEFINITIONS

We all evaluate all we do in all aspects of life. Sometimes, this is intuitive such as simply mulling over a challenge or sometimes more structured: "How did that recipe work? What would make it better? What did the people who cooked, ate or paid for the meal think? What will I do differently next time?"[1]

Within education, the term tends to be reserved for when this is done systematically in a way that can be said to produce a result (box 1). This should not distract from valuing or more importantly, seeking to improve, the day to day sometimes tacit evaluation that is part of being a clinician and a teacher. Indeed, the much praised ideal of the reflective practitioner describes one who moves from evaluation and reflection before or after action to doing reflection-in-action, that is, continuously, often subconsciously, evaluating and adapting what he/she does while he/she is doing it.[2]

EDUCATIONAL EVALUATION AND PAEDIATRICS

Paediatric learning is challenging at all levels. We want our students, trainees and colleagues to have a firm grasp of an increasingly wide range of care spanning critical emergencies, inpatient care for the sickest, ambulatory care, and community and long-term care, with an overarching consideration of the family and social aspects of children's lives. We need to do this with a dwindling number of inpatients due to intended shifts in where care is carried out, with shorter admissions and reduction in secondary care follow-up in outpatients.

Overall, we see less of the critical emergencies we want learners to be exposed to, have less children "around" for teaching and learning and often more complex children in secondary care than in previous times. Also, learners are in posts for a shorter duration and for fewer hours when they are there.

There is much to celebrate here. Children are healthier and in hospital less. Our trainees are less likely to be making errors while tired and unsupported out of hours.

Box 1 Definitions

▸ Evaluate: "To "reckon up", ascertain the amount of; to express in terms of something already known" (Oxford English Dictionary Online[3])
▸ "Evaluation is systematic determination of merit, worth, and significance of something or someone. Evaluation often is used to characterise and appraise subjects of interest in a wide range of human enterprises, including the arts, criminal justice, foundations and non-profit organisations, government, health care, and other human services." (Wikipedia[4])
▸ "Educational evaluation is the systematic appraisal of the quality of teaching and learning." (Wilkes and Bligh[5])

The teaching challenge is to maximise the available learning in the most appropriate setting, be that inpatient, outpatient, community, primary care or beyond healthcare systems. We need to know whether this new type of learning is utilising opportunities well and achieving the learning we intended. All of this needs good evaluation.

EDUCATIONAL CONCEPTS BEHIND MEANINGFUL EVALUATION

We are all accustomed to evaluation forms at meetings and those who teach in formal courses will be well used to receiving a "learner evaluation" after the session, perhaps in the form of scores, graphs or "free text" comments from feedback forms. Something is clearly being recorded and the information handed on, but what is being evaluated and does it improve anything?

The classic models of evaluation are the evaluation cycle (figure 1) and Kirkpatrick's hierarchy of levels of evaluation[6] (figure 2) and various adaptations and refinements of this model.[7] They are worth considering in detail, but in essence, they show that evaluation is about change and therefore is cyclical. Also, Kirkpatrick's hierarchy shows that evaluation often occurs at a very basic level. This is perhaps as a result of a further key principle – the tendency to measure what is

measurable (who turned up, did they think it was useful?) rather than what arguably matters more: Did the session change anything? Did the next case of disease x get managed better? Did the new knowledge improve the patient's outcome?

EVALUATION CYCLE

At its simplest level, evaluation is central in any planned learning activity. The faculty, teacher or learner defines a learning need and plans how this is going to be met in a curriculum or individual session. Once the session has been planned, it is delivered and data (ranging from simple or formal observation, written feedback, ongoing outcome measure, etc) are collected and, vitally, reflected on and analysed. This feeds back into planning as it is key to realise nothing is ever perfect and that evaluation is a continual cycle. As noted by Morrison, "If the results of an evaluation show that no further development is needed, doubt is cast on the methods of evaluation or the interpretation of the results".[8]

The experienced teacher will realise, even at the planning stage, that the key question after the session will be whether the identified need was met and therefore will already be considering how to evaluate the activity prior to undertaking it. This should not detract from the activity, however. There is always a risk of measuring the measurable rather than the important and a real risk of changing a desired educational activity to make it easier to evaluate and to therefore prove an outcome has been achieved. If external pressures are such that the overall aim of the teaching event becomes the achieving of good evaluation scores, it is potentially far easier to prove effect by rote learning items of knowledge than by trying to achieve more challenging outcomes. Teachers should therefore be aware of the need to evaluate what they do but be wary of the evaluation process dictating the curriculum.

The described evaluation cycles may seem very formal, but it is actually what many good teachers do instinctively – plan what they are going to do, do it, collect information about how it went (often simply the teacher's own impressions and learner reactions) and reflect on what will need to change for next time. Formalising the process helps to make it constructive, avoids the self-perception that everything is alright or that it has failed completely and helps teachers to respond to the requirement of external stakeholders to know that what was done was worthwhile.

Figure 1 Evaluation cycle.

KIRKPATRICK'S LEVELS

Educational outcomes can be measured at a variety of levels from the most basic (did the learners like the session?) to the more complex. This is classically shown using Kirkpatrick's hierarchy.[6] Interestingly, this is now usually represented as a pyramid[10] despite Kirkpatrick himself never using this illustration. The source of this adaptation is obscure but may represent confusion
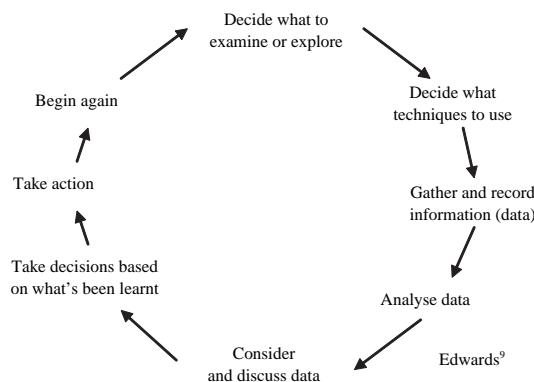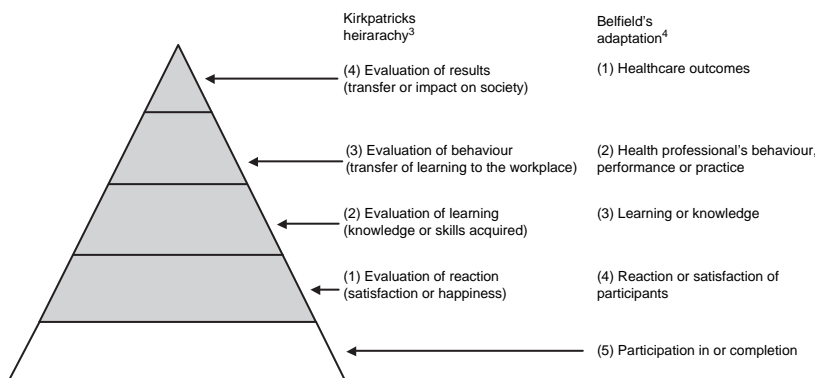

Figure 2 Kirkpatrick's levels.

with Miller's pyramid of assessment of clinical competence.[11] This shape is interesting whatever the origin. It underlines that most evaluation is done at the lower levels and that at the peak of evaluation is the impact on society (in medical terms, often patient outcomes). It shows that it is desirable to climb the pyramid and evaluate at a higher level. Also, it should be recognised that the base underpins the higher levels and therefore, if you do not achieve learner satisfaction, you are unlikely to achieve higher levels of outcome.

This base level should not be confused with giving learners exactly what they want. Learners may only have the insight into what they need once they have learnt it or indeed sometime later when they put it into practice. This is particularly true of inexperienced learners who will often give high satisfaction marks to didactic sessions based on listing knowledge not realising that this is unlikely to be remembered well.

Kirkpatrick's hierarchy has been further adapted for medical learning, particularly for assessment of best evidence in medical education by Belfield and colleagues.[7] Interestingly for medical use, they had to squeeze in a whole new level below the base of Kirkpatrick's hierarchy. This allowed categorisation of the most common evaluation undertaken in medical education and historically in continued professional development recording – that the participants had simply turned up. This is a striking illustration of how far we have to go in evaluation of medical education.

MEASURING THE MEASURABLE
One of the likely reasons for this low level focus on whether participants attended, irrespective of any educational gain, is that it is easy to measure, as is, up to a point, getting them to fill in "feedback" forms. As noted, however, just because something can be recorded does not mean it measures what someone might think it measures or what needs to be measured.

This is a key challenge in evaluation. Questionnaires are easy to distribute and collect. Unfortunately, they only have answers to questions that the questionnaire writer has thought of and they need to be carefully constructed. Interviews and focus groups are flexible and can explore areas that might not have occurred to the evaluator. They can be time consuming and doing them well is a skill in itself. Observation of sessions can provide valuable information but needs to be done carefully. There are many good general sources for those wishing to get started on evaluation[12] including guidance on questionnaire writing. There are also many written observation tools available on line for a variety of purposes and useful sources to help the interested begin.[13] Analysing interviews, focus groups and self-reflection is itself a challenge. How will you know that you have not simply seen what you wanted to see and emphasised the issues you feel important rather than what is actually going on? This brings

all the challenges that those who have delved into qualitative research will have come across: "What methods of analysis can we use that are practical, communicable and non-self deluding – in short, will get us knowledge that others can rely on?"[14] This may appear to be overcomplicating things by requiring teachers to delve into whole areas of unfamiliar theory and techniques. This is true, up to a point, although the alternative is to fall back on very basic statistics (attendance or crude measures of "happiness") that tell us very little.

Finally, evaluations that link specific teaching to "real world" (ie, the workplace) outcomes are few and far between. This may be because they are challenging. Linking a particular teaching intervention or strategy with an improvement in patient care is particularly difficult because the intervention (the teaching event) is rarely the only thing going on that might affect the outcome. Additionally, most published interventions are very brief and often of optional "add ons" to courses rather than curricula or full courses.[15]

POLITICS
An accepted truth in education research outside the medical or healthcare arena is that all teaching and research is in some sense political. This is not a concept clinicians are used to, although given that much empirical research is funded by those most interested in a particular outcome (drug manufacturers), perhaps this is something we need to be more aware of.

Evaluation is, however, one area of learning and teaching where the political aspects are more clear cut. Indeed in the USA, the term evaluation is much more explicitly fiscally linked, with one definition of an evaluator being "one who assesses a programme of research, education, etc, esp. as regards cost-effectiveness".[3]

In the local and national political sense, the key questions are who is evaluating what and why? Is this a locally conceived, owned and largely formative process designed to improve things for their own sake or is this part of an attempt for a funding or certifying authority to force specific change (such as the teaching of particular material or teaching in a particular way) or to justify shifts in posts and resources? Indeed, some feel that we increasingly live in an evaluation society with organisations being required to justify what they do and the way that they do it. Indeed, all clinicians are likely to have to provide evaluation of their teaching as part of revalidation.

This is ironic, however, compared to changes in national policies which some claim are increasingly not evaluated in an appropriate timescale.[16] Indeed in the education field, Wall notes, "The direct use of evaluation evidence in educational policy decisions is unusual".[17]

A FINAL THOUGHT
More profoundly, measuring something does not intrinsically improve it: "Weighing a pig does not

make it fatter".[18] Indeed, some would argue that a quality improvement system cannot simply feed back information but must be undertaken in a way that intrinsically improves quality.[19]

GOOD EVALUATION IN PRACTICE

Given the complexities already noted, there is a risk of not bothering with evaluation; life is busy, evaluation is time consuming and often of low fidelity and good evaluation requires entirely new ways of thinking. Newcomers should not be deterred, however. The greatest gains in quality are often made by relatively simple changes. The simple act of deciding you want to improve and evaluating and planning some action as a result is often the start of a beneficial circle with improved learning and teaching resulting in more teacher satisfaction and more motivation to improve further.

As noted, the key starting point is clarifying what the activity is that needs evaluation and why. What purposes will the evaluation serve? Is this about local improvement, curriculum change, justifying funding, etc? Is it already known how to best teach a skill and a unit simply wants to check teachers are using best practice? This is akin to an educational audit. Is the agenda changing a programme and wanting to check whether it is better? In this case the focus will be looking at the desired outcomes of the learning. Are more fundamental questions being asked about the appropriateness of the learning objectives themselves? This requires evaluation of high-level outcomes such as what learners actually achieve in the real world.

Answering such questions will give a feel for the type of data needed, which level of Kirkpatrick's hierarchy needs addressing and how detailed and exacting the evaluation needs to be. Finally, it is worth thinking through what will be done with the outcomes. Who do you need to convince? What quality of output will this require? What about confidentiality of teachers or learners?

WORKED EXAMPLES
I have tried to improve the learning experience on ward rounds. How will I know how this is going?

Clinical teachers are adapting what they do all the time. Sometimes this can be intuitive. For example, a consultant teacher feels that the ward doctors are uncomfortable that they do not get to discuss the cases that interest them on the ward round. The consultant decides to ensure there are 5 min discussions on a topic of their choice on every round and overall, everyone seems happier. At some level, this ongoing evaluation/action sequence is immensely powerful. The teacher responsible sees an issue, acts quickly and rapidly sees the results. A more formal evaluation is necessary when the problem seems challenging or complex, where others need convincing or simply as an exercise in ensuring a teacher remembers to do it. In this case, the teacher may seek verbal comments, individually or via a brief focus group, hand out feedback questionnaires, ask a colleague to observe and feedback on this section of the ward round, etc. The key then is to remember to do something with that data and to complete the cycle by then making changes.

"Evaluation" is required, what do I do?

The key question is who is requiring evaluation and why? What is the agenda of the organisation? Are they looking for proof of money well spent? If so, they are likely to have a fixed idea of what constitutes appropriate evaluation and what does not. Also, the evaluation needs to refer to the required outcomes of the organisation (such as reduction in a particular type of risky event, widening access to a particular learning experience, ensuring particular groups of learners are shown to be competent in particular things). If they are more interested in developing what is happening and seeing what is possible rather than measuring it against a standard, there may be much more flexibility. This should not be seen as the easier option. It is often more straightforward to satisfy an organisation that wants specific boxes ticked than one looking for less specific types of evidence.

Our students need to see patients in a different setting. How will we know if this is working?

Change is often driven by external factors. When this happens, the temptation is to evaluate the new (eg, teaching a group of students basic paediatric history skills in a primary care setting) against the old (eg, doing so on the ward with inpatients). This is likely to favour the old. There will always be teething problems with doing something differently. Learners are often change adverse, perhaps feeling they are being used as guinea pigs and may worry that they will do less well in exams, so measures of learner satisfaction may be misleading.

A more balanced approach is to go back to the original learning outcomes and evaluate old and new against these. For example, the advantages of a captive inpatient population may be outweighed by an increasingly inappropriate case mix. The techniques used for this evaluation will then depend on the scenario, importance, time available, etc. Observational techniques may work well early on, along with flexible techniques such as focus groups, allowing rapid acquisition of data, institution of changes and rapid re-evaluation. Evaluation of long-term impact may need measures of success, such as checks of skill acquisition or student examination results and indeed having "old" and "new" systems in operation together would allow comparative work more akin to an intervention study. This produces parallel courses, however, and the lack of equity in learning experience, particularly in high stakes situations, may make learners anxious.

My "evaluation" was emailed to me this morning, what now?

When "evaluations" arrive, there are a number of possible responses. Many scan through to see if their scores are "OK" and therefore that they are not in trouble. Some will instantly compare to any scores provided for the peer group ("at least I'm not the worst"). Some will focus entirely on the positive ("I scored well on x despite everything else") or the negative (focusing on a few derogatory free text comments rather than the graphs that show that overall the individual is rated highly). Feedback in the form of scores or free text from learners is only "data", however, and a well-designed evaluation system will ask the teacher to do something with this data and to feedback on actions. Even if this is not required, it is a useful developmental exercise. It is valuable to see that scores are good, but if you are a skilled teacher, there are probably things that could go better. If the scores are low, there is clearly work to be done, but the organisation sending out the feedback may have advice or resources. Even if there is a punitive element they will be far more impressed with an individual who has taken on board their strengths and weaknesses and has a plan than someone either in denial (blaming the system or the learners) or seeming unable to use the data to make things better.

TAKING THE NEXT STEP

For those taking on significant educational roles, there are a few final areas to consider. There are many ways of evaluating and there are good sources from within medicine. These include Wall's overview of evaluation[17] and Fleming's guide to specific techniques,[13] both prepared for the Association for the Study of Medical Education. It is also worth remembering that most of the work in educational evaluation is from beyond medicine and some will wish to find out what the rest of higher education is up to,[20] or even delve into the experience in industry. For pointers as to where to look, the UK Evaluation Society web site is invaluable (http://www.evaluation.org.uk/).

Those wanting to evaluate at a high level of learning outcomes, particularly for external stakeholders, will need to become comfortable with the language and the practice of educational research. This is not exclusively concerning flexible techniques rather than intervention studies, but is often analysed with an appreciation of concepts often new to those from a realist tradition. Those who are struggling with papers mentioning ontology, epistemology and described political stances should start with straightforward guides such as the *British Medical Journal*'s recent series on qualitative research[21] or Robson's excellent overview.[12] This is also an excellent source for issues such as questionnaire design. Those with more confidence would gain much by studying the contrast between the intuitive work of Stake[22]

and the more structured work of Huberman and Miles.[23]

These sources are best approached with an open mind. Medical research comes largely from a particular tradition of thought, that is, that reality is single and fixed and is fully discoverable by comparative experiments. Other ways of viewing the world and particularly the complexities of the social world can seem quirky and can initially seem to lack validity. Educational pedagogy can often appear in this light.

REFERENCES

1. Evaluation Trust. **What is evaluation**? See http://www.evaluationtrust.org/evaluation/evaluate (accessed 7 November 2008).
2. **Schön DA**. *The reflective practitioner: how professionals think in action*. Aldershot: Ashgate, 1991.
3. **Oxford English dictionary online.** See http://dictionary.oed.com/entrance.dtl. (Accessed 1 January 2010)
4. **Evaluation**. See http://en.wikipedia.org/wiki/Evaluation (accessed 5 November 2008).
5. **Wilkes M**, Bligh J. Evaluating educational interventions. *BMJ* 1999;**318**:1269–72.
6. **Kirkpatrick DL**. Evaluation. In: Craig RL, Bittel LR, eds. *Training and development handbook*. New York: McGraw-Hill, 1967:87–112.
7. **Belfield C**, Thomas H, Bullock A, *et al.* Measuring effectiveness for best evidence medical education: a discussion. *Med Teach* 2001;**23**:164–70.
8. **Morrison J**. ABC of learning and teaching in medicine: evaluation. *BMJ* 2003;**326**:385–7.
9. **Edwards J**. *Evaluation in adult and further education: a practical handbook for teachers and organisers*. Liverpool: Workers' Educational Association, 1991.
10. **Hutchinson L**. Evaluating and researching the effectiveness of educational interventions. *BMJ* 1999;**318**:1267–9.
11. **Miller GE**. The assessment of clinical skills/competence/performance. *Acad Med* 1990;**65**:S63–7.
12. **Robson C**. *Real world research: a resource for social scientists and practitioner-researchers*. Oxford: Blackwell, 2002.
13. **Fleming WG**. *The observation of educational events*. 2nd edn. Edinburgh: Association for the Study of Medical Education, 1998:33.
14. **Miles MB**, Huberman AM. *Qualitative data analysis: an expanded sourcebook*. 2nd edn. Thousand Oaks, CA: Sage, 1994.
15. **Zwarenstein M**, Reeves S, Barr H, *et al.* Interprofessional education: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2001;(1):CD002213.
16. **Pollock AM**, Godden S. Independent sector treatment centres: evidence so far. *BMJ* 2008;**336**:421–4.
17. **Wall DS**. *Evaluation: improving practice, influencing policy*. In: *Understanding medical education*. Edinburgh: Association for the Study of Medical Education, 2007:41.
18. **Wilkinson PR**. *Thesaurus of traditional English metaphors*. London: Routledge, 2002.
19. **Denning BW**. *Corporate planning: selected concepts*. London: McGraw-Hill, 1971:373.
20. **Popham WJ**. *Educational evaluation*. 2nd edn. Englewood Cliffs, NJ: Prentice-Hall, 1988.
21. **Kuper A**, Reeves S, Levinson W. An introduction to reading and appraising qualitative research. *BMJ* 2008;**337**:a288.
22. **Stake RE**. *The art of case study research*. Thousand Oaks, CA: Sage, 1995.
23. **Huberman AM**, Miles MB. *The qualitative researcher's companion*. Thousand Oaks, CA: Sage, 2002.