The Open University

Open Research Online

Citation

Oliver, Martin; Harvey, Jen; Conole, Grainne and Jones, Ann (2007). Evaluation. In: Conole, Grainne and Oliver, Martin eds. Contemporary perspectives in e-learning research; themes, methods and impact on practice. Open and flexible learning series. Abingdon, Oxon: RoutledgeFalmer, pp. 203–216.

URL

https://oro.open.ac.uk/15277/

License

None Specified

Policy

This document has been downloaded from Open Research Online, The Open University's repository of research publications. This version is being made available in accordance with Open Research Online policies available from Open Research Online (ORO) Policies

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding

Chapter Fourteen - Evaluation[1]

Martin Oliver, Jen Harvey, Gráinne Conole, Ann Jones

The importance of evaluation has grown in recent years so that this topic has become the focus of considerable policy and research interest (Oliver, 2000). As new learning technologies emerge there is a need to evaluate how these are used to support an increasingly diverse student population. All staff are now expected to carry out evaluations to account for resources or to justify strategic initiatives. Additionally, for individuals trying to instigate change, evaluation data can be important in providing relevant information to initiate, support and empower change by the "production of knowledge that makes a difference" (Patton, 1997).

Evaluation is not a simple, standardised practice; it has evolved to meet the needs of many different groups. Consequently, it can be hard to define, although a common definition involves both describing and judging in terms of both merit and worth (Guba and Lincoln, 1981).

Furthermore the relationship between evaluation and research more generally remains contested. Evaluation can, in fact, contribute to research as well as providing feedback for a changing teaching and learning practice. Both processes may use the same methods and study the same things. However, one way to distinguish them is to consider how findings are used. If they are interpreted by an immediate, local audience and used to support decision making, the study was probably an evaluation; if findings are interpreted in terms of theories and are presented as a contribution to knowledge, it was probably research. This emphasis on judging may be typical, but it is not universal. Not all evaluators feel that evaluation should both describe and judge; those who position themselves in the ethnographic tradition, for example, argue that their work should be non-judgemental, concentrating instead on providing credible and plausible accounts of observed practice (Jones, 1998a).

Values

The concepts introduced above need explanation. As described by Guba and Lincoln (1981), 'merit' and 'worth' are conceived of as part of a broader concept of 'value'. 'Merit' refers to intrinsic qualities, whilst 'worth' refers to extrinsic or contextual value. These definitions are rendered problematic by the critiques discussed in chapter two, and the shift in knowledge production away from positivism. The alternative to talking about these as intrinsic properties is to treat value and worth as socially constructed. This renders all aspects of an object's value extrinsic and contextual. For this reason, it makes sense to talk in a less precise but more intuitive way about things having worth if their value can be determined in terms of input/output efficiency (e.g. if they are financially effective) and merit if their value arises from moral or philosophical positions (e.g. if a particular initiative encourages qualities that society things are 'good', such as reflection, whether or not it is cost effective).

---

[1] 5768

This perspective highlights the political nature of evaluation, and calls into question whether evaluators can ever be objective. While studies might seek to help individuals by developing shared understanding of how the world works, information can also be used by policy makers to control processes and make the world work more effectively. There are however clear links between these two processes: both are empirical, leading to the construction of models and judgements about practice (Oliver, 2000).

Finally, although the process may be value-driven, it is possible to re-frame evaluation so that it is not primarily judgmental, but instead focuses upon helping people to see the values in something, enabling them to become judges for themselves: "If connoisseurship is the art of appreciation, criticism is the art of disclosure" (Eisner, 1998).

Changing approaches and different schools

Broadly, approaches in evaluation range from positivist approaches focussed upon objective data collection (typically using quantitative methods) to interpretivist ones more rooted in constructivism (typically using qualitative methodologies). Guba and Lincoln (1981) chart a brief history of educational evaluation from the use of achievement tests in the 1890's onwards. Initially, 'evaluation' was a subsidiary concept to measurement and the emphasis was on the scientific testing of individual differences using standardised tests, often with little reference to the curriculum (or wider context) from which the students were drawn.

However, in the early 1960's, educational evaluators began expressing dissatisfaction with this approach, arguing that studies needed support refinement and improvement, not simply provide judgements that endorsed or condemned courses. There was an important shift during the 1970s with the rejection of the established evaluation format (which was described rather scathingly as the 'agricultural-botany paradigm', reflecting the idea that curricula were 'applied to' students like chemical fertilizers on plant crops) as being unable to meet its own criteria of objectivity or to help educational practitioners. Importantly, this shift in emphasis led to the distinction between *formative* (shaping) and *summative* (judgemental) evaluations still recognised today.

Interest currently lies in the development of systems which minimise additional data collection but which still provide adequate information to make informed judgements about an institution's procedures. Unlike action research, in which judgements arise from iterative cycles of study and reflection of local contexts, the models underpinning quality assurance are typically imposed, and this can limit the usefulness of the study for understanding or improving things (Chelimsky, 1997). Although quality systems within UK Higher Education are now moving away from the data-heavy processes of the 1990s, the principle of accountability that resulted in the introduction of such systems remains strong, both in the UK and America (Davies iet al., 2000; Feuer *et al.*, 2002). In America, these calls have already been followed through, with the 'No Child Left Behind' act advocating in law that educational research should use quantitative approaches to implement evidence-based practice (Feuer *et al*, 2002).

The rise of a new approach did not lead to the old being abandoned; indeed, both remain visible in educational research today, and the controversies over method and meaning that led to the split remain unresolved (Hammersley, 1997). There have been numerous attempts to unify the two schools; although none has been successful, these attempts have given rise to yet other approaches. Patton (1997), for example, has suggested that the argument between the two traditions has distracted evaluators from a more important issue: that no matter how closely a study adheres to the principles of its tradition, if the report that is produced sits on a shelf unread, then the investigation was pointless.

These controversies make it clear that methods are not neutral; there is no single approach that all evaluators would agree is 'good'. Moreover, any method can be used inappropriately. As a result, it is important to understanding the limitations of each approach in order to ensure that any conclusions that are drawn are warranted.

A series of methods currently used to evaluate e-learning is discussed below.

Experimental methods

Within the experimental tradition, studies are often based upon comparisons between conditions in order to demonstrate improvement or identify differences. Studies also rely on the ability to control the 'noise' introduced by the complexity found in the classroom so as to focus on the effects of just a handful of variables. Educational resources can be trialled in this way as part of the design process, especially when determining usability issues. Early evaluations also attempted to use this approach to demonstrate the effectiveness of computer-based approaches (Gunn, 1999). There was an emphasis on quantitative data collections, benchmarked ratings of courseware or the production of statistics to show one method was better or equivalent to another (e.g. that software was 'better' in some absolute sense than a chapter of a book; see the WinEcon evaluation – Allen *et al*, 1996 – for example).

However, numerous researchers have argued that hypothesis testing was simply not the best way to understand the complex real-world situations where the materials would be used which, unlike laboratory settings, were not amenable to control and causality was difficult to demonstrate (see, e.g., Oliver and Conole, 1998). The use of a control group is also ethically problematic (denying access to potentially beneficial opportunities), and intervention studies introduce issues of authenticity and transferability of findings. Contextual elements (such as access to supportive peers, parents or paid tutors) were seen to affect performance in unexpected but powerful ways; changes to computing access policies, teaching staff or resources and infrastructure can all change the nature of an innovation as it is studied; students who experience bad teaching often work harder as a result, hiding possible negative effects; and, of course, educational contexts are social settings, so there is often nothing to stop those with access to an innovation sharing with those who do not. Unpredicted but beneficial outcomes emerge that are not part of the study. Experiments can only happen when the research process interferes with this social setting (for example, by taking people out of it and placing them in a lab) – although experimental researchers do recognise that the relationship between their findings and classroom practice ('ecological validity') may be problematic.

Illuminative evaluation

Illuminative methodologies developed in response to experimental approaches by adopting ethnographic strategies. This provided a more open-ended, exploratory approach to evaluation (Parlett and Hamilton, 1972), emphasising the need to follow up unanticipated developments, to show an awareness of the influence of context (which traditional methods sought to eradicate through controlled experiments) and thus to 'illuminate' the focus of the study. Evaluators took a 'neutral outsider' stance in order to observe practices as they happened. Adopting such an approach was only possible on a small scale, which contrasted with the apparent need for increasingly larger scale experimental methods, but also made this process both time intensive and costly. Such a 'value free' approach sometimes results in a mismatch between the sponsor's expectations and study outcomes.

Systems approaches

The systems approach model emerged during the 70's as a way of providing evaluation feedback linked to learning outcomes. This approach also aimed to move away from scientific methods, as formative feedback linked to learning would help inform ongoing judgements as well as contributing to a body of knowledge about education and training (Hamblin, 1974). Many studies are intended to improve students' learning, even if only indirectly.

Such an outcomes-focused approach was sometimes criticised as being mechanistic and limited, and almost positivist in style. There were also issues about whose objectives were being addressed and how this information would be fed back to policy makers. There is also a perceived risk – shared with quality assurance systems – that asking questions about achieving particular goals changes practice rather than describing it (Blalock, 1999).

Goal-free evaluation

Rather than focusing upon stated intentions and to help eliminate the bias of stakeholders, goal-free evaluation (Scriven, 1972) takes a more constructivist approach, pursuing issues that emerge as having significance rather than pre-determining the study focus. Potentially, this provides a broader perspective, exploring unanticipated outcomes through working and spending time with the various stakeholder groups. This approach also differs from the research-like illuminative approach in that its purpose is to help make informed decisions about practice (rather than just describe it) through its interventionist use of dialogue between stakeholders.

Action research

The rise of action research models marked an important trend in educational evaluation, producing a new body of educational research knowledge as well as influencing evaluation methodologies. Action research has its roots in the work of the psychologist Lewin who sought to understand and model how individuals acted in society (Lewin, 1952). He emphasised the importance of working with the people being studied in order to effect change, an approach that resonated with the problems facing teachers who sought to evaluate what took place in their own classrooms (McNiff, 1988). Action research has developed into a movement within education that

places value on the ideas of democracy, emancipation and collaboration; within action research, it is argued that evaluation should not be something done to subjects, but something done *with* participants (Kemmis, 1996). The evaluation process thus develops iteratively: it moves through cycles of planning, acting, observing and reflecting which build upon each other, sometimes progressing (for example, by developing and refining a model of classroom practice, or taking an increasingly detailed focus on a particular issue), sometimes diverging (for example, by highlighting an important but previously neglected issue that the evaluator considers worthy of investigation), depending on the interests and values of those involved (McNiff, 1988).

However, it should be noted that – in spite of their similarities, such as a concern with generating locally meaningful understanding – there are important differences between action research and evaluation. Evaluation does not always share the democratising agenda; action research may be relatively unconcerned with judging; and action researchers usually study their own practices whilst evaluators typically study other peoples'.

Responsive and utilisation–focussed evaluation schools

Approaches that adopt a pragmatic focus towards educational evaluation. These involve tailoring an evaluation so that it will "make a difference" (Patton, 1997), which inevitably involves taking into account the viewpoints of various stakeholder groups. The popularity of this method reflects evaluators who seek to act as change agents; using evaluation helps initiate this change.

Different approaches utilise different ways of integrating stakeholder involvement. Stake's responsive evaluation (1980), for example, structures activities in a way that is responsive to the various stakeholder groups' needs. Guba and Lincoln's fourth-generation evaluation approach (1981) seeks to identify stakeholder group's concerns and negotiate a strategy before data are collected, thereby taking into account a range of different values and perspectives.

Patton (1997) starts from the premise that a good evaluation is one that enables people to do things. Usefulness is thus privileged over rigour, and any method can be used so long as the audience of the report will find it credible. The importance of rigour is not ignored, however; utilization-focused evaluators still aspire towards it. The focus on use means that evaluations have to be timely ('quick and dirty' is preferred to 'perfect but late') and informative – and importantly, they must inform someone who can act. In contrast to the democratising principles of action research, say, this approach emphasises rhetorically persuasion, with powerful stakeholders as an audience. The utlization-focused school of evaluation can be seen as a good example of ways in which processes have become commodified. Because this approach is organised around the notion that studies should be 'consumed', it can simply serve those who have power (for example, those who can afford to commission studies). As a result, it is important that utlization-focused evaluators remain sensitive to issues of audience and politics so that they can take a radical position rather than reinforcing inequality (cf. Freire, 1993).

Integrative evaluation

The integrative evaluation model was proposed by Draper *et al.* (1994) as a way to evaluate resources within a broader education context by placing the audience and their needs above methodology. It proposes the use of a barrage of data collection techniques to provide a rounded picture of students' and teachers' practice with technology. Issues and anomalies can then be remedied, until the use of technology becomes seen as a normal and natural part of the course.

Evidence-based practice

Whilst these different schools of thought seem to imply that educational evaluation has moved steadily away from its roots in experimentation, this is not entirely true. Particular forms of evaluation rooted in quantitative methods and aspiring to scientific status have gained prominence in the current era of educational accountability, even though it has been noted that policy-based, outcomes-driven audits of practice run the risk of distorting rather than reporting the practice they seek to describe (Blalock, 1999). Additionally, there have been calls from policy makers in the UK to adopt the idea of evidence-based practice developed in medicine for use in education (Davies *et al.*, 2000; Fitz-Gibbon, 2000). In America, these calls have led to the 'No Child Left Behind' act, which advocates in law that educational research should implement evidence-based practice by using quantitative approaches (Feuer *et al.*, 2002), ignoring widespread criticisms that the approach is methodologically and ethically inappropriate (e.g. Hammersley, 2001; Oliver and Conole, 2003; St. Pierre, 2002).

As part of the medical model, an hierarchy of evidence has been drawn up that states explicitly the degree of faith placed in different kinds of study. Randomised control trials, following the classic experimental model, sit at the top of the hierarchy as a gold standard, with qualitative methods (such as all those described above) relegated to the level of anecdotal evidence, either discounted entirely or at best given a role in supporting the findings from favoured types of study. As a reaction against 'unsystematic' research, it appears that educational evaluation has come full circle.

The ethics and politics of evaluation

The process of evaluation cannot be separated from "issues of power, politics, value judgements and human interests" (Esterby-Smith, 1994). Evaluation is intended to improve, make changes or allocate resources. As a consequence, the communication and negotiation process is one that needs to be handled carefully.

Research into ethical issues in the evaluation of learning technology is not common, but should underpin evaluation. Cohen and Manion (1994) provide a thorough guide to the ethical issues associated with educational research; these are echoed in the guides to ethical conducted produced by groups such as the British Educational Research Association (BERA, n.d.). These include, for example: getting access and gaining acceptance; privacy; anonymity; confidentiality; and deception. The cornerstone of addressing these issues, argue Cohen and Manion (1994), is informed consent. This is the principle that each individual has the right to freedom and self-determination; participants in research should thus have the right to understand what they are taking part in and to end their involvement at any point. However, as they point out, there are many problems with this principle: in some studies where

deception is necessary (such as studies involving covert observation), providing a full explanation is impractical; there is an inevitable tension between providing an audience with enough information about research participants and ensuring that they cannot be identified; there are also many debates about whether the rights of individuals can ever be compromised to further the good of the many. Because these tensions are complex, they must be interpreted on a case-by-case basis. As a result it is important that educational research studies discuss these issues; in particularly complex cases, the whole purpose of a paper might be to explore such issues as a case study.

Issues in evaluating learning technology

The messiness of e-learning

As described in chapter one, there is considerable confusion about what 'counts' as e-learning, and also what terms should be used to describe it. This is a particular problem for evaluation. Many of the approaches described above emphasise the need for dialogue between stakeholders; in e-learning, this may prove particularly difficult since each may have different – even contradictory – assumptions about what e-learning is and how it works. This makes the initial scoping phase of work particularly important.

Doing evaluation

A recurrent issue in e-learning is that practitioners are expected to evaluate their own initiatives – even though academics have rarely received training to do this, see no particular benefits and have other priorities (Harvey *et al.*, 2002). As a consequence, efforts have been made to support academics with guidance, tools and models.

A number of evaluation frameworks have been developed that articulate the process of evaluating e-learning, such as the CIAO! framework (Jones *et al.*, 1996), the integrative approach (Draper *et al.*, 1994), the SECAL method (Gunn, 1997) and the ELT framework (Oliver *et al.*, 2002). Conole and Oliver (2002), for example, have attempted to model the process of evaluation based on studies of practice. This model splits the process into a sequence of steps. Of course, this is an idealised representation of practice; in reality, progress through these is unlikely to be linear, and not all studies or evaluators will follow every step. By starting from studies of practice, this research assumes that evaluation is a socially constructed practice. An alternative approach would be to work out the 'essence' that defines 'evaluation' as a concept, perhaps through philosophical enquiry.

According to this model, evaluation consists of the following stages:

1. Identification of the audience(s) for the evaluation
2. Selection of an evaluation question
3. Choice of an evaluation methodology
4. Choice of data collection methods
5. Choice of data analysis methods

6. Selection of the most appropriate format(s) for reporting the findings to the audience

These steps can be thought of as a combination of contextual (1, 2 and 6) and mechanical (3, 4 and 5), or alternatively as strategic and tactical choices. A similar distinction is advocated by Draper *et al* (1994).

Placing the audience and their needs above methodology locates such models firmly within the school of utilization-focused evaluation; in other traditions, the methodology would influence the kinds of questions that were considered to be acceptable – so, for example, experimental studies would require well-defined hypotheses, whereas illuminative studies would typically involve open-ended questions.

Tools to help with evaluation

A number of projects have used the kinds of models outlined above to developed tools for novice evaluators. Ehrmann (1999), for example, complemented a model of the process with a data bank of survey questions. Other researchers have pursued more flexible developments – for example, the online evaluation toolkit (Conole and Oliver, 2002). Such developments use the model to form an interactive resource that prompts reflection, records decisions and provides advice on the options available at each point in the process. However, to date, the evidence that these tools change practice is limited. They have been shown to change evaluation *plans*, but the longitudinal studies needed to demonstrate an impact on practice are lacking (Oliver *et al.*, 2002).

A number of other peculiarities influence data collection around e-learning. Much useful data can be captured, including transcriptions of online dialogues. Nonetheless, researchers must be cautious not to place too much faith in such records, since they only portray part of the picture – the social contexts that surround such data may be completely hidden (Jones, 1998b). Further complications arise from the distance that often exists between evaluator and participants: methods such as interviews, which rely on face-to-face contact, may be impractical where participants are geographically dispersed, and will need to be rethought if conducted using new technologies since participants' responses differ (Oliver, 2001). The same is even true for online surveys, which tend to attract responses from different groups than paper-based surveys, changing the impression given by the data.

Issues of method

Although whole texts have been written on issues of methods, there are several particular difficulties for e-learning evaluation. One recurrent issue, for example, concerns drawing comparisons. Experimental studies rely on the ability to draw comparisons between conditions – however, since educational innovations such as the introduction of technology may change the nature of what is learnt, any comparison between this and previous practice must be viewed with scepticism. It has been argued (Oliver and Conole, 1998), however, that actually this happens all the time; what is important is not whether or not comparisons can be drawn (you could ask, for example, "how do these two groups perform on this exam?"), but what they *mean* (for example, "have these two groups learnt different things?"). Such a change of emphasis echoes practices within qualitative traditions of research, where case studies

are important. Although these are the study of particular cases, either authors or readers often feel the need to compare what is reported with other reports or with personal experience (Stake, 1994). Here, questions are not about whether such comparisons are valid but whether they are educative. Thus the issue of comparisons now covers a range of emphases, from the technical problems of valid inference about models (or the world) to a social or psychological focus on meaning-making.

Another issue facing evaluation is the question of authenticity. It is well established that people involved in studies perform differently from those who are not. This phenomenon is known as the Hawthorne effect. It is also recognised that the presence of an observer (for example, in a classroom) changes the way in which people interact. For studies in education, these problems are extremely important because they imply that the process of doing research changes what it is you are researching (Kvale, 1996). Some traditions have embraced this issue – it is, for example, the whole point of action research that such studies *should* change, not just comment upon, the world (Kemmis, 1996). Other evaluators try to overcome this issue by trying to blend in (for example, following the ethnographic tradition; cf. Jones, 1998b) or by seeking to study day-to-day activities in an unobtrusive manner for at least some small part of their process (for example, the CIAO! framework; Jones *et al*, 1996).

Causes are also hard to determine within education, since contextual elements (such as access to supportive peers, parents or paid tutors) can effect performance in unexpected but powerful ways; changes to computing access policies, teaching staff or resources and infrastructure can all change the nature of an innovation as it is studied; students who experience bad teaching often work harder as a result, hiding possible negative effects; and, of course, educational contexts are social settings, so there is often nothing to stop those with access to an innovation sharing with those who do not (Oliver and Harvey, 2002). This has implications for experiments, which are only possible *because* the research process interferes with this social setting (for example, by taking people out of it and placing them in a lab), inevitably causing problems for the relationship between their findings and classroom practice ('ecological validity').

Any study is only as good as the data that it draws upon (Patton, 1997). This observation is particularly important in the context of studies involving children, although some researchers would argue that the same concerns hold true for data concerning any individual's experience of education. In addition to pragmatic problems such as participants who seek to please the researcher (telling them what they want to hear) or subvert the process in some way, there is also the educational problem that people cannot always articulate what they have learnt. Some knowledge, it is argued, is tacit; although we have learnt something, we cannot put it into words (McMahon, 2000). Similarly, we may have learnt how to do something, but this is no guarantee that we will choose to do it (Barnett, 1997). Thus studies can only provide partial accounts; as a consequence, it is important for the evaluator to appreciate what the limits on their data are.

With any study, irrespective of methodology, the question of how widely the conclusions can be generalised must be considered. Again, qualitative methods often make this concern explicit, valuing 'rich' descriptions of contexts so that the reader can decide upon the extent to which this case resembles their own (Stake, 1994). In

experimental studies, however, the reader must infer from any demographic information given whether or not the participants in the research are in any way similar to the pupils who might be involved in some learning development. This demographic data represents a model of what the researcher believes is significant: for example, that the important variation between individuals can be described in terms of classification according to age and gender. Since students vary from year to year, institution to institution (let alone country to country), as well as in terms of what might be described as their social class, a model such as this can easily be argued to neglect important variables. Even the decade in which the study took place has been shown to have an important influence on the design and interpretation of studies (Berliner, 2002). If the model is incomplete in these potentially important ways, then the confidence with which conclusions can be generalised must be called into doubt. And, of course, all models are incomplete; they are selective representations, emphasising some features and ignoring others. This makes it important to recognise the rhetorical way in which they operate. Indeed, it has been argued that rather than trusting the general conclusions of experiments over the provisional conclusions of case studies, experiments ought to be doubted and questioned in exactly the same way as the tentative findings of qualitative studies (Holt and Oliver, 2002).

There are ethical as well as technical issues associated with comparative evaluations. Can one group be denied access to a new form of education – particularly if the researchers suspect it might be more effective and hence give an advantage on formally assessed work? One solution to this problem involves piloting initiatives within non-assessed areas of the curriculum. This remains problematic, however, because assessment has such a profound effect on how pupils act (Biggs, 1999); the situation may thus change dramatically once assessment is re-introduced. Other alternatives include crossover designs, in which the group experiencing the innovation swaps with the group who do not have access to it after some mid-point measurement or observation, or simply providing access to the resources for revision purposes once the study is complete.

The role of the evaluator

The role of the evaluator is one that necessarily changes with each different methodology. For some this might mean acting as a neutral outsider or administrator of predefined tests while for others this involves becoming a 'critical friend', educator, advocate or even lobbyist (see e.g., Patton, 1997). As the level of participation increases within the process, so their role becomes more influential in determining both evaluation process and product. Besides methodological and technical competency based on their training in systematic inquiry and analysis, evaluators are likely to need skills in communication and team building, group process and negotiation (Guba and Lincoln 1989).

This may imply that the evaluator is (initially) separate to the stakeholders, being brought in to undertake an evaluation study. This might be possible in funded projects but for day-to-day practice it is likely that one of the key stakeholders will have to take on the role of evaluator. Both options introduce issues. How will personal interest and investment influence data collection methodologies? How easily can the roles of (say) assessor and evaluator be moved between? By comparison, which factors will influence the selection of an external evaluator, and how much of an

impact does their possession of 'evaluation skills' or personal goals have upon the resultant process?

The influence of context

Earlier, the process of evaluation was described as starting from an awareness of different audiences and their concerns. Designing studies around the interests of such audiences positions evaluation as being a practical and political activity. As Weiss (1993) argues, "evaluation is a rational enterprise that takes place in a political context", which "by its nature makes implicit political statements about such issues as the problematic nature of some programs". Unlike illuminative or experimental studies, this kind of evaluation cannot be a search for 'truth' in some positivist, revealed sense. Instead, it is a process that seeks to inform and educate in order to inform subsequent action (Patton, 1997). This raises important ethical questions for evaluators. If evaluation enables action, whose actions will be supported? Whose agendas will be served? Who is paying for the evaluation?

> The concept of educational evaluation, like the studies it involves, critically depends on context, and some contexts are more regulated than others. Regulation can provide useful guidelines or a stick for stakeholders' backs. The key is to offer tangible benefits to learners, teachers and institutions, preferably in that order. Prescribed methods are not always conducive to this aim, as common [ab]use of student evaluation of teaching questionnaires shows. Used well they provide one source of feedback to guide improvement, used badly they become an obstacle course around exam time and a whip to spur teachers on through the tenure track. A truly authentic evaluation involves collaboration among teachers, developers and learners using existing knowledge, experience and impact analysis, to create quality learning opportunities in a given context. Knowledge of established theory and methodology may be limited, but the 'common sense' 'grass roots' 'design-based research' approach still works well. It may not produce generalizable results, but then neither do many other 'approved' methods.
> *Cathy Gunn*

Traditionally, those who are served are those who commission the study; the people in the study are positioned as research subjects and treated as sources of data. Other forms of evaluation (such as action research or utilization-focused evaluation) have tried to subvert this situation, treating people as research participants or collaborators, rather than subjects to be 'treated'. In some cases, the process of evaluation alters to reflect this new mindset. Participants can re-design the study; they can conduct their own, independent enquiries that are incorporated into the design; they may provide points of view not invited by the study; they can be invited to discuss the agenda of the study and the appropriateness of the methods used; and they may be reporting the findings directly to the funders rather than having their stories re-interpreted by evaluators. Whether any of this is a useful or important thing to do will depend on the specific situation being evaluated. What is important is that the evaluator should remain aware of the different political agendas that their study might be used to support so that they can take a principled position about which groups' agendas they will work to support (Oliver and Harvey, 2002).

Broad stakeholder involvement can make the evaluation process extremely complex. To address this, some approaches use the negotiation of concerns between evaluator and stakeholder groups as part of the process (e.g. Guba and Lincoln, 1989). The process of clarifying the purpose of an evaluation or confronting the data generated can help stakeholders to appreciate each others' relative values (Eisner, 1998) as well as realise the educational value inherent in that experience (Patton, 1997). Conflict

between stakeholder groups may also result, however, in which case the evaluator must act as a facilitator, negotiating outcomes and the implications of decision-making. In order to fulfil such a facilitative role effectively involves gaining stakeholders' cooperation and trust then sustaining their interest and involvement over an extended period of time (Guba and Lincoln, 1989). However, such involvement is not without its cost: the evaluator's perceived loss of 'objectivity' may result in reduced trust or value from other quarters; this can reduce the impact of their findings or hinder access to sources of quality data since they are no longer viewed as neutral.

Conclusions

Evaluation may be seen as an increasingly important part of educational practice, but the research effort that has been invested in this topic has served to complicate, not simplify, its practice. Evaluation remains problematic and contested conceptually, technically and philosophically. Within such a complex area, it becomes increasingly important for individual researchers and practitioners to be clear about their assumptions and their theoretical commitments.

Evaluation also serves to illustrate the politics that surround e-learning. The explicit consideration of stakeholders and their agendas foregrounds the issues that can arise. Provision of evidence may serve the interests of sponsors or it may be challenging, radical or educative. Stylistic decisions about whether a study is an intervention, a judgement or a description reveal how evaluation plays a part in this book's themes of interactivity and of change. Fundamentally, the idea that evaluations should lead to use – a tenet of utilization-focused approaches – illustrates the process of the commodification of knowledge.

Because of these complications, it has become impossible to advocate 'good practice' in any simple or generic way (although the development of models and toolkits may allow practitioners to follow accepted practice amongst the community of evaluators). Raising awareness of these issues amongst researchers and practitioners is thus increasingly important; this, together with efforts to understand the process and significance of evaluation in ever greater depth, represent important directions for future research in this area.