Towards effective practitioner evaluation: an exploration of issues relating to skills, motivation and evidence

Jen Harvey
Dublin Institute of Technology
Dublin, Ireland

Martin Oliver
University College London
London, UK

Janice Smith
University of London External Programme
University of London, London, UK

ABSTRACT
Although academics are increasingly expected to undertake studies of their practice, particularly where this involves the use of learning technology, experience to date suggests that meeting this expectation has proved difficult. This paper attempts to explain this difficulty. After reviewing literature that provides a rationale for practitioner evaluation, the experiences of three projects (EFFECONDS, ASTER and SoURCE) which attempted to draw on this process are described. Three main areas of difficulty are discussed: the skills and motivations of the academics involved, and the kinds of evidence (and its analysis) that ‘count’ for a given evaluation. This discussion leads to the identification of a number of problems that inhibit practitioner evaluation, including ambiguity in the nature and purpose of evaluation, and a general feeling that the function of evaluation has already been served through existing quality mechanisms. Finally, the possible implications are considered of some or all of the steps in the evaluation process being undertaken by an evaluator working alongside the academic.

Keywords
Evaluation, learning technology, academic roles, participant evaluation

Introduction
Individual academics’ involvement in educational evaluation has changed in recent years, from being passive recipients of evaluation findings to being expected to become an instigator or active participant within the evaluation process (Oliver, 2001). This has arisen, in part, through the changing roles and responsibilities of academic staff as well as ‘creeping managerialism’ (Holley and Oliver, 2000) as institutions become more accountable in their provision of courses. The need to cater for diverse student populations in a variety of ways or, in this case, to satisfy stringent requirements for project funding, are also drivers behind this new evaluation imperative.

In parallel, emergent methods of learning delivery - particularly electronic environments - are giving rise to sophisticated new research areas. By necessity, therefore, academic staff are now expected to engage in effective evaluation practice, to ‘prove’ the worth of online forms of delivery. To support this activity, a range of resources and course materials, including evaluation tools and toolkits (e.g. Conole, Crewe, Oliver and Harvey, 2001), cookbooks (e.g. Harvey, 1998) and targeted staff development, are now more widely available. Such materials and training can provide guidelines and assistance in evaluation planning and design.

This paper will build on the experience within the Effective Framework for Embedding C&ITs (Communication and Information Technologies) through Targeted Support (EFFECONDS) project, for which evaluation was a key learning outcome. The project’s outcomes suggest that, even when explicit evaluation resources and support are available, few academics engage with the process, and many of those who do, do so superficially. This conclusion is also supported by evidence drawn from the ASTER (Assisting Small-group Teaching through use of Electronic Resources) and SoURCE (Software Use, Re-use and Customisation in Education) TLTP projects.

The experience gained through these projects highlights lessons learned in encouraging effective evaluation of learning technologies, enabling identification of a number of issues relating to skills development, staff
motivation and the collection of evidence of effective evaluation practice. The paper concludes with the identification of some of the implications of this review for researchers and evaluators working with academics.

**Background**

Why is practitioner evaluation important? As well as being perceived as a necessary part of the implementation process, effective evaluation strategies have been shown to impact on the resultant success of centrally funded projects. In a study of 104 university IT projects in Australia, for example, some of the reasons identified for government-funded projects not attaining the desired learning outcomes were ‘not evaluating the project in the anticipated context of use’ and conducting ‘limited or poor evaluation of the project’ (Wills, 1998).

For those involved in carrying out an evaluation, the processes required (for example, clarifying aims and objectives and negotiating strategy with relevant stakeholders) can become a ‘change inducing experience’ (Patton, 1997) with the potential for developing effective evaluation practice in order to enhance the student learning experience. In an evaluation of the ASTER project, Davis, Oliver and Smith (2001) identified a need for lecturers to develop a more critical awareness of their own skills and through the process of reflection (Schon, 1987) to move towards improving their practice. Involvement in the discussion of practice can also facilitate individual and group development, and be in itself of educational value through the shared exploration of personal experience (Beetham, 2001). In addition, the creation of communities of practice based on such knowledge resources, e.g. evaluation experience, has been put forward as a method to facilitate the way in which novices and practitioners can develop and disseminate effective practice (Saunders, 2000).

Thus in addition to the usual extrinsic motivations for evaluation, such as the improvement of systems through feedback or the production of evidence in support of judgements of worth, it can be argued that evaluation should also have an intrinsic educational value. In the next section, project work which attempted to draw on these various motivations in order to encourage practitioner evaluation will be described.

**Context**

This consideration of evaluation skills and issues has arisen, in part, from the experiences of the EFFECTS project and the finding from the evaluation studies (Harvey and Oliver, 2001). One of the aims of this project was to embed learning technologies into 70 mainstream teaching modules. This was to be achieved by developing a nationally-accredited programme of professional development in embedding learning technologies (www.seda.demon.org/pdaf-elt/), offering the programme to interested staff and supporting them through the process.

In response to funders’ requirements, the project had, from the start, a well-developed evaluation strategy (Oliver, Conole, Phelps, Maier, Wilkinson and Bailey, 1999). The multi-level evaluation plan comprised three layers, utilising participant evaluations, as illustrated in figure 1.

**Figure 1:** The evaluation structure for the EFFECTS project
The three-level structure was based on the assumption that each new level would synthesise the findings of the level below, and would add new topics for evaluation that were of relevance to stakeholders involved at this level. So, for example, the institutional teams would synthesise the case studies produced by participants on their courses, and in addition, would address issues of the effectiveness (including cost effectiveness) of their programmes.

By the end of the project, only half of the expected number of case studies had been produced, and the quality of these varied considerably. Although other outputs far exceeded expectations (see appendices to the project’s final report; Beetham and Bailey, 2001). However the external evaluation report (Harvey and Oliver, 2001) notes the difficulties participants had in achieving the evaluation outcome.

The ASTER and SoURCE project evaluations have generated similar evidence. Interestingly, all three TLTP projects supported and made explicit a common philosophy: that practitioner evaluation of practice was valuable and to be encouraged for the reasons cited earlier. But all three of the projects had problems implementing and attaining this objective. The next part of the paper will attempt to explore some of the issues that arose.

**Issues facing practitioner evaluation**

The experiences of the three projects described above highlighted a range of problems that limited academics’ engagement with the process of evaluation. These issues have been grouped into three broad areas – skills, motivation and evidence – each of which will be considered in turn.

**Skills**

Evidence from both the EFFECTS project final report (Beetham and Bailey, 2001) and the external evaluation (Harvey and Oliver, 2001) suggest that skills are an important factor in influencing whether or not academic staff undertake evaluation studies. Some staff simply seemed to lack the specialist skills required, which suggests a need for training and support in this area. Other staff, however, appeared to have the skills but to lack the confidence to carry out a study. It is less clear in such cases how support can best be offered.

Although evaluation skills are not easily acquired or commonly practiced, as mentioned previously, a number of toolkits and wizards have been produced to help academic staff in the evaluation process (Conole et al, 2001). As part of the EFFECTS external evaluation process (Harvey and Oliver, 2001) the level and nature of support for each of the seven outcomes was explored. As the project developed, levels of support changed and areas had been prioritised for additional support. One such area, in some institutions, was evaluation. The University of North London, for example drew on local evaluation expertise and associated developments (Media Advisor, see www.unl.ac.uk/ltri/demo/) to help with participant evaluation strategies.

Similarly, the ASTER project employed reflective tools (http://cti-psy.york.ac.uk/aster) to assist academics wanting to employ communications and information technology (C&IT) supported learning, whilst SoURCE (http://cortez.open.ac.uk/source/) offered a fee for those contributing case studies with evaluative elements. Despite different approaches - ASTER team members co-wrote case studies and SoURCE paid for them - neither project was any more successful than EFFECTS in obtaining evaluation studies. Whilst this points to the deficiencies of a number of simultaneous centrally-funded projects utilising the same methodology, it also reveals a lack of good evaluation practice within a limited community.

The issue of staff support and development is, however, more complex than it initially appears. Findings from these studies also suggested that none of the methods employed to support these participants were sufficient to encourage them to undertake robust evaluations. Only practitioners who had gone through the process a number of times, and perhaps only through this iterative process, seemed to have developed the confidence or skill to undertake evaluation at any meaningful level. Perhaps, as Tickner (2000) has suggested, a combination of fixed resources and self-study resources, such as those used to support EFFECTS programmes, are not sufficient to replace face-to-face support to a diverse range of participants. Instead, it may be that developing the ability to evaluate is more about learning to participate in the practices of expert evaluators rather than simply acquiring a set of skills (Brown et al, 1989), and that close personal support coupled with experience is required before academics feel ready to carry out studies autonomously.
Although it is important to understand whether academics have the skills required to evaluate learning technology, it must be recognised that simply being skilled is not enough. As Barnett (1994) has noted, individuals may choose not to practice their skills for any one of a variety of reasons. Thus whilst skill may be described as being a necessary condition for practitioner evaluation, it is not sufficient to cause such activity: what is also required is the disposition to use these skills. In the next section, issues of motivation, which provide insight into the disposition of academics towards evaluation, will be discussed.

**Motivation**

The lack of case studies, and the poor quality of some of the case studies that were produced, appeared to be an indicator that evaluation was either not taking place or was only occurring at a superficial level. As noted above, even when given support to develop an awareness of the need for evaluation, this was still insufficient to ‘cause’ academics to evaluate.

This raises a fundamental question for projects that value practitioner evaluation. If staff were prepared to spend the time undertaking a full EFFECTS programme, for example, why did they not see evaluation as being an important part of this process? Worryingly, the EFFECTS external evaluation (Harvey and Oliver, 2001) identifies many academics who openly question, ‘Why evaluate?’ Out of the 78 participants involved in this evaluation study very few had engaged or had planned an evaluation strategy for their implementation. Perhaps most tellingly, 30% commented on a question asking about evaluation strategy, saying that the question was not applicable to them. This was felt by tutors to be, in part, a result of the timing of the questionnaire survey carried out by the evaluation team. The reasons why participants might be motivated to undertake programmes like EFFECTS is the basis for another study (Smith, 2001).

There did not appear to be any trends in the subject backgrounds of participants that might account for this, but there was some evidence that those registering were not new members of staff. This would suggest that these individuals would have brought with them some teaching expertise but not necessarily in the use of learning technologies. By registering for accredited M-level modules, participants were clearly demonstrating an interest in new ways of learning and teaching with technology. It may be that such programmes can, in time, increase exposure to practitioner evaluation through the provision of tools and techniques that may otherwise remain outside an individual academic’s normal practice.

Many EFFECTS participants experienced problems in collecting evidence to meet the evaluation outcome. At UNL, for example, a variety of causes were cited: from a lack of enthusiasm for another layer of evaluation that was not valued by the institution, to a lack of confidence in what could or should be measured. Extrinsic forms of evaluation, such as that required by project funders and TQA assessments (subject review) clearly made participants ‘evaluation weary’. Not knowing what - or how - to ask is evidenced by case studies weak in evaluation evidence. One participant commented wryly that students ‘didn’t give me any feedback’.

It was only through the application of short-term solutions to their educational concerns, and the problems that resulted from these, that staff began to recognise the value of developing a strategy and carrying out an evaluation study. With hindsight, this is unsurprising. It is only when the direct relevance of evaluation as a way of understanding problems of personal importance becomes apparent to them that time and effort will be allocated to this, rather than to the other pressing duties that fill academics’ time (cf. Smith & Oliver, 2000).

If there is no perceived problem with a method or a ‘gut feeling’ suggests things are going well, why look to change things? The case of one participant who tried to instigate a web-based evaluation (minimising participation overheads for students) of a web-based resource complained of a lack of feedback from the students. In this case, a poor response to an evaluation study cannot be seen as helping to improve teaching practice which is then compounded by the relatively low value institutions generally place on such concerns (Warren, 2001). Warren identifies a ‘lack of recognition or reward for outputs such as papers or educational journals’, as well as a lack of time more generally, as being possible causes for the low-level of evaluation resulting from his work at Southampton.

Given a lack of disposition to consider issues relating to strategy and integration, it is therefore of importance to consider ways in which academic staff might begin to see the value of carrying out an evaluation at an earlier stage. This issue will be returned to later in this paper.
Evidence

The previous sections have outlined some of the reasons why academics might not start to undertake practitioner evaluation. However, it is important to realise that the problems continue even after the process has begun. One fundamental problem that arose in all three of the projects described was what ‘counts’ as evaluation.

In education, evaluation serves a complex function, often attempting to meet the needs of several stakeholders simultaneously. However, a closer analysis of interests in evaluation revealed that different groups wanted very different things from these studies, leading to ambiguity and conflict over the purpose of evaluation.

- The academics wanted to gain experience – however, few felt there would be any direct benefit to them in writing this up (Smith & Oliver, 2000, Beetham et al., 2001).
- A number of institutional programmes involved central funding and the institutions and funders want quality measurement and accountability rather than deeper understanding.
- The educational researchers want to understand the impact on learning – but there are competing views of learning, and no study can cover them all. Irrespective of which view of learning is adopted, all approaches require a more complex (and thus time intensive) study than that requested by the funders, and with doubtful rewards (Beetham et al., 2001).

The implication of this is that evaluation becomes contested – pulled between formal and informal studies – and between audit and research methods. In such contested situations, it is perhaps unsurprising that academics have difficulty engaging with evaluation.

Most of those responding to the EFFECTS study relied either on informal or existing methods, such as end of module forms rather than methodologies designed for their projects, even if EFFECTS tutors had provided consultancy support. The ASTER project evaluation study obtained similar evidence, describing evaluation methodologies as being superficial, with staff tending to use a course questionnaire. ‘These typically involved a set selection of general questions, many of which fail to engage in any deep or theoretical exploration of the course, and which are often only superficially analysed’ (Davis et al., 2001). Institutional methods, e.g. curriculum reviews or internal audits, are often perceived as adequate for obtaining feedback and, in most cases, are mandatory. With limited time, it is likely that staff will spend this time developing the skills involved in using learning technologies rather than developing the new skills involved in evaluating them, given that they may well perceive that institutional mechanisms already exist for this process.

If academic staff did not perceive evaluation as a method to improve practice or they were finding that their limited experience in the evaluation process raised more issues than solutions, then it is not surprising that they did not write up their work. This lack of understanding of the purpose of evaluation could perhaps have led to the mismatch between their reliance in using more intuitive informal methods to derive feedback, and the researchers’ and staff developers’ expectation that they would produce educational research case study evidence.

The case studies that were completed by participants also varied in quality. Within EFFECTS, there was initially resistance to producing a case study template but, to support participants, a template was devised based on the seven learning outcomes. Other projects such as ASTER and SoURCE had also used a case study format to capture practitioner practice (ASTER, 2001; SoURCE, 2001). (It should be noted that there was no agreed template for the case studies across the three projects.) However, these methods were designed to try and encourage participants to become reflective practitioners as well as forming part of the dissemination and project development strategy. Despite devising templates with pointers to issues to be covered, the case studies still varied in quality and could not be considered exemplar practitioner evaluations in themselves. Some project websites (see http://sh.plym.ac.uk/eds/elt and www.unl.ac.uk/ltc/effects) collated evaluation resources including articles and checklists and encouraged participants to access them for evaluation support.

Across the projects, one solution to the shortage of case studies being produced involved using support staff to assist academics in the evaluation process. Most commonly, this support entailed talking with the academic and then writing a short (c. 1,000 words) synopsis of their account of the study. This pragmatic solution clearly reflects the differing needs and interests of both lecturers and educational researchers in the evaluation process, but it does raise questions about whether or not these studies can be considered to be practitioner evaluations. The ASTER evaluation report (Davis et al., 2001) questioned the reliability of the data collected in the cases studies, commenting on the superficiality of approach; and latterly, discursive records of practice were utilised to capture implementation practice more effectively (ASTER, 2000).
Discussion

Underlying the entire EFFECTS approach was the assumption that it would be beneficial to get academics to evaluate their own practice. It was felt these evaluations should be written up as cases studies, or as part of the portfolio or disseminated during seminars. This was to provide evidence of evaluation. However, in light of the issues raised above, it is important to question this assumption. Why is it important to have practitioner evaluation? In addition to the educative value to participants outlined earlier, a number of other motives arose from the projects under consideration:

- It privileges the voice of experience/authentic perspective
- It may provide access to otherwise unobtainable data
- It is less intrusive, so more ‘authentic’
- It is cheap – salaries of evaluators are already covered
- It enhances academics’ practice by encouraging them to become more reflective practitioners

Given that practitioner evaluation proved difficult to achieve, it is also worth considering what might be lost if evaluators provided help with some or all of the steps of the evaluation process.

- With evaluation design – no obvious loss, although the designer will impose an ideological framework that might distort subsequent data collection and analysis (cf. Barnett, 1994).
- With data collection – intrusion into an authentic situation that may give rise to a Hawthorne effect
- With data analysis – the interpretation of data inherently reflects the perspective of the evaluator; thus intervention may lose the perspective of the practitioner
- With data presentation – this may limit the expression of personality of the academic
- Throughout the process – may not enhance the academic’s practice

Given that the most effective solutions adopted in the projects involved some or all of the evaluation effort being taken away from the practitioners, it is necessary to reflect carefully on which, if any, of these are important to preserve. This will allow greater support to be provided, whilst maintaining those elements of most interest to the community.

Whether online or face-to-face, it was clear that groups of academics found value in being able to discuss the issues that they encountered in their use of learning technology. Such discussions have a considerable educational value; these groups clearly resemble the dialogic educational practices advocated by Lipman (1991), or the inquiry-based programme of professional development described by Rowland (2000). Thus while these groups cannot be forced, or caused to happen, they do seem to offer a valuable way of sustaining, encouraging and developing the kinds of benefits advocated for practitioner-based evaluation at the outset of this paper.

Although many of the participants mentioned that working with like-minded people was the best thing about being involved with EFFECTS, this kind of support was not as highly rated in terms of relative usefulness to their projects. In some of the interviews, participants commented about the missed opportunity for the use of groups’ collective expertise within EFFECTS. However, Southampton and Plymouth were active in trying to set up and maintain groups with common interests. Through these group meetings, staff were encouraged to present their evaluation findings as part of their project.

I don't think we had enough opportunities to share what we'd been doing. I don't know enough about other people to know what they'd been doing. I know I went to one session, probably a one day session, where we listened to and experienced other people's projects but I wasn't aware of there being more opportunities like that.

I think it would have been really nice if we'd all known; we could have talked through it as we were going through our portfolio and the other thing is that we could have then provided evidence for each other which would have given it a much wider framework.

Another feature of the EFFECTS programme was to set up and maintain online discussion groups. These were used to complement face-to-face activities, to encourage peer support and encouragement and to provide experience of working online. However, providing these opportunities did not lead to them being widely used. Perhaps there did not appear to be a perceived need to become involved if face-to-face sessions were also timetabled, particularly if online activities were not a direct requirement of the course. But again, by focusing discussion, whether online or face-to-face, around the various stages of the evaluation process and encouraging participants to explore evaluation data, then perhaps this activity might become more mutually beneficial and thus subsequently increase in importance.
The inclusion of such opportunities for dissemination does, however, expose academics to scrutiny by peer reviewers. People are discomfited if evaluation data collected is negative or identifies what might be considered to be areas of difficulty, such as the data collection failure noted above. This may lead to participants reporting only positive information for fear of peer criticism. Such sessions then need to be carefully managed in order to achieve their intended outcomes and become more effective in the design and refinement of the evaluation strategy.

Conclusions

The experiences described in the projects above suggest that there are a number of unresolved problems facing practitioner evaluation. Not least amongst these is the ambiguity in the expectations of researchers and evaluators in the area. For a professional evaluator, such ambiguity is potentially an asset, since it allows flexibility for the final report to be shaped according to their expert judgement in order to be of greatest use to the intended audience (cf. Patton, 1997). However, for the novice, it appears to become a problem of under-specification, making their role all the harder to complete. Faced with freedom of choice over evaluation methods, for example, novices may well opt for familiar approaches rather than those best suited to the particular task in hand.

Ambiguity is particularly prevalent in the terminology used in this area. A ‘case study’, for example, can imply any one of a number of formats or methods depending on its context. This is a consequence of the fact that case studies are more properly an approach to evaluation, rather than a method; the evaluative techniques used to implement the study can and should be decided once the focus of the study has been chosen (Denscombe, 1998). Even when well-specified and clearly understood, academics may lack the motivation to carry out studies of learning technology. Unless the academic already has an intrinsic interest in educational research in their discipline, such studies bring relatively few rewards; similar amounts of effort, invested in a more familiar field of study, are likely to generate more directly relevant outputs such as mainstream disciplinary publications.

These motivational concerns call into question the very notion of practitioner evaluation in this context. If it is unreasonable to expect academics to spontaneously carry out evaluations themselves, we must then ask how much of this process – if any – they will be willing to engage with. Thought must be given to the reasons why practitioner evaluation is considered important; it may well be that, with adequate resourcing, their workload can be reduced without losing the qualities that make this kind of study valuable, or else their workload could be reduced in other areas to enable them to do the evaluation.

Whilst practitioner evaluation is important, achieving this remains problematic. Evaluators and researchers in this area must begin to develop a sense of when such an approach is appropriate, what it is reasonable to expect different academics to contribute, and what support can be provided without rendering the process meaningless. If this can be achieved, and academics engaged in a more moderate way, then there is every reason to hope that through this involvement there will be an increase in interest which then motivates the subsequent development of practitioner’s skills.

References

ASTER TLTP project http://cti-psy.york.ac.uk/aster


