Evaluating online teaching and learning

Martin Oliver

University College London, Gower Street, London, WC1E 6BT, UK
Tel: +44 20 7679 1905; Fax: +44 20 7679 1715; E-mail: Martin.Oliver@ucl.ac.uk

Abstract. Evaluation is becoming increasingly important, both as a part of the design of online courses and as a mechanism for quality assurance. In this paper, the issues facing the evaluation of online teaching and learning are considered. Different motivations for evaluation are identified, and strategies for addressing these needs are illustrated.

1. Introduction

The drive for transparency and public accountability in the UK’s public sector has had a far-reaching impact on Higher Education. Part of this impact has been an increased emphasis on evaluation [1]. However, the drive to evaluate has not been matched by support and training for the practitioners who are supposed to carry out these processes (see, e.g., [2]).

In response to this, several initiatives have been implemented to provide practitioners with support, such as the development of toolkits [3], cookbooks [4] or manuals of advice and guidance [5]. What many of these resources lack, however, is specific advice on evaluating online learning and teaching. Consequently, this article will include a review of the issues specific to this topic, supplemented by illustrative cases. To structure this review, however, it is first necessary to elaborate the reasons for evaluating online learning and teaching.

2. Background

2.1. The characteristics of distance learning and of online learning and teaching

As noted above, much of the literature addresses generic issues of evaluating learning technology rather than concentrating on the particular characteristics of online learning and teaching. However, it is important to take these into account when designing and implementing an evaluation. In order to do this, the characteristics of distance learning will be outlined, and then refined to consider the special case of online learning and teaching.

As a useful first step, Peters [6] identifies several distinctions between distance learning and traditional forms of study. These include:

- A shift from an elitist model to mass higher education;
- A move towards increasingly structured and planned programmes of study,
- The industrialisation of course development, including the division of labour amongst teams of specialists;
- The challenge of maintaining dialogue as a central component of distance courses;
- The loss of informal opportunities for learning, for example in social settings.
Importantly, although physical distance is taken into account in his analysis, Peters is primarily concerned with "distance pedagogics". Under this view, techniques such as broadcast lectures, because they add nothing to traditional forms of teaching, are grouped with traditional teaching techniques. By contrast, open learning self-instructional materials are considered alongside distance learning techniques, even when the specific packs are distributed to campus-based students. In this paper, however, physical distance will be considered, since it introduces a range of pragmatic problems for evaluation.

Online learning and teaching is harder to characterise, due to its relatively short history and diverse forms of implementation. Clearly, most online programmes will mirror the characterisation of distance learning provided above. Importantly, however, several new characteristics may also be added. These include:

- Technical requirements;
- Skills requirements;
- The breadth and use of different media.

These characteristics are equally important for staff and students involved in online learning. (See, e.g., [7].)

Another important aspect of online learning and teaching that is worth noting is that many systems automatically log use. Particularly important is the way in which online discussion – compared closely to traditional correspondence learning by Peters [6] – is recorded in full. As will be seen, this is of considerable use to evaluators.

2.2. Evaluating learning technology

The term "evaluation" refers to a wide-ranging collection of methodologies. It is used to cover review processes such as checklists as well as empirical judgements. Furthermore, the term is sometimes confused with assessment [5]; however, in this article, it is only taken to cover processes that support judgements of value and worth of programmes [8].

Reviews of evaluation methodologies have stressed the importance of determining the purpose that the process will serve. Numerous distinctions have been made, but the following set (from [9]) has been adopted as a useful summary within the learning technology community:

- Formative evaluation;
- Summative evaluation;
- Illuminative evaluation;
- Integrative evaluation;
- Evaluation for Quality Assurance.

These five purposes will be used to structure the following discussion. In addition, special attention has been given in the literature to the evaluation of costs and to comparative evaluations. As they represent special cases, these two categories will also be considered. However, although costs will be discussed as a self-contained topic, comparative evaluations will be considered in the context of summative evaluation, which is where they most frequently occur.

2.3. Summary

In this section, the scope of this review has been established. The key characteristics of online learning and teaching and reasons for evaluation have been identified. In the next section, the evaluation of online
learning and teaching will be considered, starting with general issues before moving on to consider each of the specific reasons for evaluating a programme in turn.

3. Evaluating online learning and teaching

3.1. General issues for evaluation

In the previous section, characteristics of online learning and teaching were identified. Of all of these, the most immediate impact on evaluation arises as a consequence of the physical distance involved. Even a cursory glance at lists of methods for gathering data (e.g., [10]) will reveal that most involve contact with students. Unless considerable effort and expense can be made to arrange meetings, methods such as focus groups, interviews and observation are rendered impractical in a distance learning context.

Many suggestions have been made that re-create these methods by proxy. Cousin and Deepwell [11], for example, have discussed the feasibility of virtual focus groups. They demonstrated that these can be an effective substitute for a real meeting, and offer all the benefits often advocated for computer-mediated communication such as allowing space for reflection when responding. However, they also noted several limitations, such as participants’ reluctance to contribute messages on sensitive topics and the need for a skilled facilitator of online discussions (see, e.g., [7]).

An alternative approach is to focus the evaluation on the types of data that online systems are good at gathering. Phelps and Reynolds [12], for example, combined web-based questionnaires (allowing immediate responses without the subsequent need for lengthy data entry) with system usage data such as the time and frequency of page access. This was achieved by using Javascript to create a tracking log via a CGI script on the host server each time a page was requested. These methods provided very rich data on patterns of usage and on users’ motivation and satisfaction. However, once again, the methods had their limitations. Usage logs are difficult to interpret, since they cannot reveal why a learner accessed a particular page or what they did with it once they had gained access. Whilst logs of this type are valuable and easy to collect, it remains important to triangulate them with other sources as part of the interpretative process.

One common technique adopted in courses of this type is to use online questionnaires. In the case above, however, the rate of return for the online questionnaire was low. This leaves the evaluation open to the criticism that the opinions recorded will be from a self-selecting sample, and thus unrepresentative of the wider group of users. In particular, it seems probable that less confident users of technology are those least likely to respond. Other evaluations of online learning (e.g., [13]) have complemented this method using a paper-based survey distributed to non-respondents of the online questionnaire, leading to a much better overall response rate.

Another common approach to data gathering in an online environment involves the creation of a “feedback” discussion area. Again, this offers the opportunity for continuous feedback from participants and also provides a full transcript of responses in an electronic format, ready for analysis. However, as Taylor et al. [13] note, this can open the floodgates to an unstructured wash of criticism, much of which may come from a small but vocal minority. These views, which may be unrepresentative, can cause considerable problems if designers feel that they need to take them all into account. Since it is impossible to please all of the learners all of the time, the value of feedback forums may be greatest when an evaluator is able to act as an intermediary between the data and the course team.

Many of these issues can be summarised by noting that evaluation in this context raises two general problems for evaluators. Firstly, with many of these methods, the process can no longer be controlled.
Opportunities for contributing data can be provided, but what the student does with this opportunity is up to them. Taking a more extreme position, it is often impossible to tell who is actually contributing the data – it would be very easy for someone else to contribute using a student’s form or login, for example. Secondly, methods for interpreting these types of data are still being developed. Whether the data be from system logs or bulletin boards, lessons are still being learnt about the most useful and appropriate ways of drawing conclusions.

Finally, it is worth identifying methods of evaluation that are not affected by the move to distance education. Essentially, these include methods designed to operate at a distance, such as postal or telephone surveys, or those that do not require empirical data. Surveys are subject to the same problems as other distance methods, such as self-selection, as discussed above. Methods such as checklists and conceptual maps, which fall into the latter category, are also subject to criticism. Whilst these are relatively easy to implement, significant questions have been raised about their value (e.g., [14]), not least because of their highly subjective nature.

3.2. Formative evaluation

Having highlighted some pragmatic problems for the evaluation of online learning, it is worth considering the purposes of evaluation. In the previous section, six distinct purposes were identified; the first of these is formative evaluation. This refers to evaluations that provide information that allows revisions and improvements to be made. Its primary audience usually consists of the project or course team.

Several features characterise formative evaluations. Firstly, they are usually carried out by a member of the project or course team; in this regard, they are “internal” evaluations. In order to be useful, they must provide timely information in a format that is readily accessible to the course team. The implication of this is that utility is likely to be a higher priority than validity for formative evaluations [15].

Within small, self-contained teams, immediately accessible evaluations (where the results need little or no subsequent analysis) such as focus groups are often useful. As noted above, however, such techniques are often inapplicable in a distance context. In addition, any source of data that relies on an input from the students will introduce delays into the evaluation when nothing can be done but to wait for responses. As a result, scheduling becomes extremely important. Ironically, although formative evaluations may be what is referred to as “quick and dirty”, they are at their most useful when carefully prepared for. The economy of effort must come in the collection and analysis of data, rather than in the planning of the study.

A good illustration of the issues involved in formative evaluation is provided by the EuroMET project [16]. This involved the development and delivery of web-based courses in Meteorology by a consortium of 22 partners. Given the complex structure of the project, it was important to ensure that appropriate information was gathered and communicated in a timely manner. The evaluation that was carried out included two strands. The first was a survey of users’ views on ease of use, pedagogy (including scientific integrity) and value as a replacement for traditional teaching methods. The second involved usability trials, carried out with a sample of five users under controlled conditions. This approach allowed the project team to identify elements of the course that worked and those that needed revision. The strong use of visual material was welcomed by users, for example, whilst inconsistencies in the material (such as variations in style and symbol use from section to section, reflecting the different contributing authors) were identified as areas that required attention. Both strands of evaluation contributed to the re-design of the system’s navigation. The evaluation showed that users found that some icons were too similar to each other, that users had no sense of where they were in the material, and that students wanted clear learning objectives and end-of-unit summaries to be added.
The timing of this evaluation allowed these points to be fed back to the project team and suitable revisions incorporated. The structure of this particular study also supported the triangulation of results, as one part of the evaluation helped to validate the revisions proposed in the first round of work.

The formative evaluation has been extremely useful in producing modules which are suitable for their target audience, are easy to use, and are instructive. The fact that the evaluation was embedded into the development work meant that it was relatively easy for the developers to modify the modules according to the recommendations of the evaluators and, in turn, for these modifications to be tested during the next evaluation phase. In particular the usability study showed that the modifications made after the first evaluation phase were effective [16, p. 192].

3.3. Summative evaluation

In contrast to formative evaluation, summative evaluation is often an external process concerned with judgement rather than improvement. It often involves assessment of a project against its aims or, in the case of online education, of a course in terms of learning outcomes.

It is often asserted that such evaluation ought to be carried out by an evaluator outside of the project team in order to assure objectivity (e.g., [17]). However, recent critiques of evaluation have made the point that evaluation is inherently political [15]; objectivity is, in many ways, a myth. Many of the proposed advantages of scientific methods, designed for use in controlled conditions, such as transferability and replicability, simply do not apply in the ill-defined, authentic world of education practice.

Such critiques have led to a division between experimental designs for summative evaluation and those that are primarily exploratory [10]. This section will focus on experimental approaches; the exploratory approaches will be explored further in the following two sections.

Experimental approaches face several challenges. One of the most significant is that it is effectively impossible to ensure control, so that factors external to the experiment do not influence the outcomes. An obvious example of a failure of control would be if an online course broke down, and students passed the final exam because they had all formed self-help groups and taught themselves from textbooks instead. Since it is impractical to control all the extraneous factors in any educational setting, particularly when it involves learning at a distance, it becomes extremely difficult to attribute causality to the teaching intervention.

More subtle problems arise in the context of comparative evaluations. This type of study is often popular with managers or funders, since it is assumed that the comparison will demonstrate whether the innovation adds any value to the learning experience. Here, however, cross-condition contamination is even harder to prevent. Even if online courses are password-protected, it is quite possible for students to share IDs, download materials or even just share notes. Even if extrinsic factors could be controlled, such issues make it hard to argue that outcomes are the result of a particular intervention.

Other problems also arise for comparative experiments. Experiments are predicated on the ability to control the context in which they take place; this is necessary in order to isolate the variables to be studied. In an educational setting, it is often impossible to do this on pragmatic and ethical grounds. If different teachers are involved, another important factor is introduced. The same is true if the materials change, the students are different, the subjects covered vary or the way in which they are taught alters. From an ethical point of view, it is difficult to justify allocating extra resources or opportunities to only a sub-set of a student body, particularly when the course carries credit towards an award.

A final criticism is aimed at comparisons of traditional and computer-based courses. This is particularly relevant for courses that are subjected to a comparison of learning outcomes “before and after”
adaptation to an online format. The argument is that, because the methods used differ so radically, these experiences are so different that they cannot be compared in any meaningful way. The analogy used is that it is like comparing apples and oranges.

It has been argued that such comparisons can be drawn, but that this must be done with care [18]. It is an easy matter to gather data on student preferences, for example, or to compare performance on an end of year exam. What must be asked, however, is what such a comparison means. If the change from traditional to distance learning (for example) really does represent a completely new educational experience, it is inappropriate for the assessment used to remain the same – a point often neglected when designing online courses. This raises serious questions about the validity of assessment methods which is beyond the scope of this paper. However, if the assessment does remain the same, then a comparison of performance – irrespective of the measure’s validity – can clearly be made. If what is being evaluated is simply student performance against some assessment yardstick, then it is appropriate to compare their net experience in reaching this. In such cases, comparison of courses becomes a viable option.

Taking into account the problems of contamination, control and transfer, what the experimental approach does allow is a firm conclusion about one particular comparison (albeit with the proviso that the measurement’s reliability should be considered critically). Claims about the generality of conclusions, however, are more difficult to justify. The implications of this are that experimental evaluations (and comparative evaluations) are possible. However, they must be designed with care, reported in a way that acknowledges the limitations of the method in an educational context, and interpreted with the same criteria as any qualitative case study.

In the context of distance education, the validity of measures must be constantly questioned since, as noted, it is all too easy to collect data that measures access rather than use, for example, or which represents the opinion of a self-selecting sample. What makes this process complex is that the evaluator may well have fewer clues as to the quality of the data as they work. In a face-to-face context, collusions may be noted, and ambiguities addressed as they arise; deprived of this context, the evaluator must work solely with the raw data. Although it may be possible to investigate specific cases, or to gain the opinion of independent observers in the field, systematic checks of the quality of the data that is received will be difficult to implement.

Hiltz et al. [19] provide a good example of an evaluation of online course that adopts a critical approach to experimental methods. The evaluation concentrates on the Virtual Classroom® system, and involved three separate studies. These considered hypotheses such as, “[Online communication and learning] can improve quality of learning as measured by grades or similar assessments of quality of student mastery of course material”. Importantly, the proviso about the validity of assessment as a measure of learning is explicitly acknowledged here.

In this report, careful attention is paid to general experimental evaluation issues, and Hiltz et al. explicitly discuss the limitations of the experimental method outside of a laboratory setting. Moreover, the extent to which results can be generalised is also explicitly discussed. For example, considering the first study of Asynchronous Learning Networks (ALNs) described in isolation, Hiltz et al. acknowledge that, “the longitudinal field study does not allow us to conclude whether better educational outcomes in ALN-supported courses are the result of collaborative learning techniques, ALN use, or both”. This problem was tackled by triangulating the three studies. This allows Hiltz et al. to conclude that, “though any one measure or method might be legitimately questioned in terms of its validity, reliability, or generalizability, the weight of several different kinds of studies over a period of five years, is convincing”.

Copyright © 2001 All Rights Reserved
3.4. Illuminative evaluation

The problems of employing experimental methods in educational settings are not new. In 1987, Parlett and Hamilton proposed an alternative model based on a “social anthropological” approach to evaluation [20]. Rather than attempting to quantify impact, these studies seek to discover the factors that are important to the participants. This is achieved through phases of observation, inquiry and explanation, with analytical methods adapted pragmatically and triangulation used to improve the reliability of findings.

In contrast to experimental evaluations, which seek to control the factors that might influence learning and teaching, illuminative evaluation seeks to describe and interpret them. The educational context becomes the focus of the study, rather than whatever measure of learning is used for assessment purposes.

This avoids many of the problems that faced experimental approaches. However, the responsive flexibility that allows illuminative evaluation to achieve this also prevents its conclusions from being objective or transferable. Conclusions are interpretations constructed by the evaluators. Confidence in them can be increased if methods are triangulated, but they remain interpretations of specific events. Although this may be perfectly adequate for summative evaluations of single programmes, it will pose problems if the lessons learnt are to be applied elsewhere. Whereas experimental evaluations make the claim that their results can be generalised, illuminative evaluations accept that transfer is problematic. Rather than attempting to deny this situation, this approach requires evaluators to engage with these problems and to interpret the findings critically to assess their credibility in each new context. In summary, illuminative evaluation accepts the criticisms levelled at experimental studies and, rather than trying to overcome them, works within the constraints that they represent.

It is the first phase of illuminative evaluation that is particularly problematic for online learning. The dispersed community of students makes it hard to know what the evaluator should observe, and even raises the question of what “counts” as observation. As noted in the previous section, it may be possible to investigate particular cases to develop a sense of what students do; however, systematic observation remains problematic. In such cases, the evaluator is faced with a choice of either making do with a selection of cases or else simply observing the “public” portion of students’ behaviour, such as performance on essays or online discussion groups.

Wegerif’s evaluation of the development of communities in online discussion provides an example of the illuminative approach to evaluation [21]. The study involved participant observation (the first phase of illuminative evaluation), in-depth interviews together with surveys and a transcript of discussion (the second phase), then analysis (the final phase). This allowed a deeper understanding to be developed of the process through which students succeeded or failed in joining an online community, and the implications of this on their achievements. The conclusions that were drawn were specific to this situation, but recommendations were put forward as a starting point for discussion – which included the suggestion that these findings be used as a point of comparison for other, similar studies.

“As well as its more specific conclusions and recommendations, this study has illustrated a method for researching the social dimension of ALNs and put forward the beginnings of a conceptual framework, including the concept of the difficult threshold between insider and outsider status, which may prove of general value in understanding the impact of the social dimension on learning on ALNs and how this impact can be taken into account in course design” [21].
3.5. Integrative evaluation

Experimental and illuminative approaches to evaluation can be seen as two extremes, each of which has limitations. Several evaluators have attempted to create compromises that incorporate elements of both approaches. Integrative evaluation is one such approach.

The term integrative evaluation is used in several contexts, but in the field of learning technology research is usually associated with the approach devised by the Teaching with Independent Learning Technologies (TILT) project [22]. This combined the structured approach of experimental evaluations with the values and flexibility of illuminative studies. In addition to the study itself, phases of work took place that addressed the context of the course, addressing issues such as policy, resources and the tacit teaching objectives of the staff involved. Integrative studies incorporate multiple methods, including within-group experimental studies of performance, surveys, interviews and confidence logs.

Inherent in the approach, however, is the assumption that the evaluation’s findings will be situationally-specific. The term “integrative” reveals the central motivation for the project team, which was to improve the way that computer-based resources (including online materials) were incorporated into the course. As with illuminative approaches, no attempt is made to generalise the findings.

Many of the methods favoured by integrative evaluation lend themselves to studies of students learning at a distance. Confidence logs, for example, involve surveying students – a method that could easily be carried out by post, albeit with reduced response rates. The same holds true for most measures of performance. If the number of interviews remains small, then these, too, should prove feasible with a distributed group of students. However, integrative evaluation was developed to provide an holistic analysis of courses; the move to the distributed, fragmented context of distance education, coupled with the reduced reliability arising from self-selecting responses and harder sampling choices for interviews, must challenge the evaluator’s confidence in the validity and completeness of their study.

Draper and Brown [23] used the integrative approach in their study of remote collaborative tutorial teaching, which formed part of the MANTCHI project. This involved around 20 different studies, each of which adopted a similar approach, and which were then summarised and synthesised in order to make claims about the project as a whole. This process allowed the evaluators to argue that the collaborative tutorial teaching processes were at least as effective as traditional methods, were received with mixed levels of enthusiasm, and were primarily of benefit in enriching the curriculum and in staff development. There was no need to generalise the conclusions beyond this point, since the approach was unique to this project; however, the synthesis of so many individual integrative studies did provide an adequate basis for summative judgement and for recommendations for others attempting to adopt a similar approach.

3.6. Evaluation for Quality Assurance

Although the concept of Quality in Higher Education is a complex one, providing evidence remains a common component for quality assurance procedures [24]. This provides a link between quality and evaluation, although one which is by no means straightforward [25]. Although quality places critical reflection high on institutional agendas, it also fosters a climate in which negative outcomes are problematic and risks are likely to be replaced by attempts at compliance [26]. In such a context, the real benefit of evaluation comes through links to quality enhancement, rather than absolute judgements. Such an approach implicitly acknowledges that failures occur, and that there will be, almost inevitably, room for improvement within any educational endeavour. The similarity between this and formative evaluation is clear.
Another role for evaluation in this context is in identifying and documenting examples of good practice. This is integral to benchmarking processes, and can help provide a basis of good examples from which standards can be derived. Again, there are clear similarities between this and the kind of synthesis of studies described above. The main difference is in the intended use of the outcomes: to further knowledge or guide practice in the case of evaluation, and to impose standards or identify room for improvement in the case of quality.

Unlike other purposes of evaluation, quality assurance usually takes place in the ‘centre’. Although there are studies that look at the experience of students as well as the service offered (e.g., [27]), most analyses of quality focus on the processes that take place amongst the service providers. The consequence of this is that although the distribution of students may change the focus of quality assurance (e.g., away from issues such as estates provision, and towards problems such as supporting students), the process through which it operates is likely to remain relatively unchanged.

Belcher et al. [28] describe a quality system for Internet ‘portals’. These subject gateways act as a ‘one-stop shop’ for users, and provide an infrastructure through which educational resources can be managed and disseminated. They serve online learning in two ways: by providing resources that can be incorporated into courses and by providing a corpus of good practice that provides academics with examples on which to base their own developments.

The importance of information management in this context places considerable importance on the quality systems for resource selection. Belcher et al. argue that ‘information gateways need to use a service-driven definition of quality, where resources are selected for their relevance to the user group as well as their inherent features’.

From this position, three roles for evaluation emerge: the identification of the information needs of users, the usability of the web-based portal site and the selection of materials. Issues relating to evaluation to explore needs and usability have been discussed above; however, the evaluation of resources poses a distinct set of issues. As noted earlier, checklists represent an evaluation method that is unaffected by the move to distance learning; however, this method is frequently criticised for being subjective (and unrepresentative) and inflexible. Belcher et al. have addressed these problems by:

- Ensuring that the staff who review resources are experts in the subject, are representative of the portal audience and are trained by existing reviewers.
- Holding “editorial meetings” where all selection staff discuss the criteria.
- Monitoring sites selected by new staff to ensure that they comply with the established criteria.
- Making the selection criteria available to users, so that the audience can develop and informed and realistic set of expectations.

Further responsiveness to quality issues can be achieved by seeking feedback from users on the service provided, so that a process of continuous quality enhancement can be implemented. In essence, this approach has taken a flawed but convenient evaluation process (checklists) and used it as the basis for an evaluation process that remains convenient but which is responsive to the satisfaction of the users of the portal.

3.7. Evaluating costs

The final use of evaluation to be considered in this article focuses on the costs of online learning and teaching. Costing any form of education, particularly that involving learning technology, is a complex process, and no definitive costing method has yet emerged [18]. One root of this problem is that costing
requires a model of the educational process, and the scope of the model that should be employed is ambiguous. There is a general consensus that evaluations of cost ought to include the development as well as maintenance costs of the resources involved. However, whether the costs incurred by students or the costs of “salvaging” resources for reuse should be considered remains less clear cut.

The responses to these problems have been varied. Oliver et al. [29], for example, advocated that costs should be gathered from sources such as accounts, time sheets and records of participation, should reflect the range of budgets that were influenced by the process (including departmental, central and those of individual students) and should also incorporate qualitative information on intangible or opportunity costs. This information should then be synthesised so that it can be used by the managers responsible for the programme being evaluated as the basis for decision making. As with illuminative evaluation, no attempt is made to produce a general result, and the emphasis is on understanding the implications of the programme.

In contrast, Bacsich et al. [30] have proposed an Activity Based Costing model for evaluations. This approach sets the model for analysis as any and all activities that are involved in supporting a programme. To this end, they have also drawn up a list of “hidden costs” – costs that are normally neglected, including institutional overheads and costs to students – although they acknowledge that no such list can be exhaustive. Importantly, the costing process still relies on centrally held accounts; the ambiguity of other costs (e.g., what an hour of a lecturer’s time is worth) are noted but not resolved.

Initially, given the emphasis on central records, it would appear that evaluating costs should also be unaffected by the shift to distance or online learning. However, it is the issue of hidden costs, central to both of the approaches outlined above, that is most affected. In the distributed context of distance learning, it becomes hard to identify the activities that students engage in. Consequently, whilst the institutional costs can be evaluated, the full impact of a programme becomes all but impossible to quantify.

Draper and Foubister [31] have evaluated the costs and benefits of the MANTCHI project, referred to above, which involved remote collaborative tutorial teaching. In their study, the hidden costs of the programmes for students were consciously omitted from the study. Instead, the study focused on the costs to the institution in terms of the impact on teaching staff.

Acknowledging the problems of accurately costing all the elements of the project, Draper and Foubister opted to retrospectively interview the staff who had taken part in the project. This method was intended to provide a clearer understanding of both the costs and the benefits of the approach; in this respect, it resembles the approach described by Oliver et al. [29] above. Lecturers’ recollections of the time required for participation were gathered and compared to existing logs of their involvement. This triangulation was used to demonstrate that, since the lecturers consistently under-estimated the time they had contributed, their account could be used as a safe minimum estimate. One point that was emphasised, however, was that a monetary value could not be substituted for these figures; what an hour was “worth” varied during the academic year, depending on the other demands upon the lecturer. In spite of the limitations imposed by the data, the study did allow an understanding of cost implications within the project. This was used to identify the situations in which the collaborative tutorial approach was most efficient in terms of saving staff time.

4. Conclusions

The increasing importance of evaluation in Higher Education raises particular problems in the area of online learning and teaching. The first of these is that methods for evaluation in this context are still
evolving. Although techniques have been proposed and validated, such as the use of systems logs, discussion transcripts and online questionnaires, the reliability of these methods remains open to question. At present, triangulation represents the best approach to ensuring the validity of the findings of studies of online learning.

More fundamental, however, is the lack of control over the evaluation process that results from working with a dispersed group of students. Many of the contextual clues vital to evaluation in traditional educational contexts are denied to the evaluator. This makes it hard to know exactly what students are doing during the course, let alone how or why they are doing this. Moreover, traditional problems such as sampling or costing are exacerbated, since many important factors will be hidden from the evaluator and will remain unidentified.

The importance of these problems varies considerably, depending on the purpose of the evaluation. With evaluation in the context of quality assurance, the distribution of students may well be a relatively small concern, since the focus will often be on activities that take place in one central location. For the summative evaluation of online programmes, however, the issues of validity and control that are raised mean that any conclusions must be interpreted with caution.

The implication of this is that evaluation in the context of online learning and teaching must involve an awareness of its limitations. The design of studies must be led by the aim of the evaluation, with ever greater levels of reflection and triangulation used to support studies where issues are likely to be missed as a result of the distribution of students. The illustrative cases described in this article have a role to play by providing examples of good practice to draw on, but as with any study, they must be interpreted with caution if the methodological conclusions are to be applied in new contexts. These cases, and the issues raised in this article, represent only the starting point for an ongoing discourse on the evaluation of online learning and teaching. The single most important conclusion to be drawn from this is not a single "right way" to evaluate online courses, but rather that by identifying some of the issues that make this area problematic, an agenda for future discussion has been set.

References


[29] M. Oliver, G. Conole and L. Bonetti, The hidden costs of change: evaluating the impact of moving to online delivery. TLTC report no. 1, University of North London, 1999
