

2002

## What does 'Impact' Mean in the Evaluation of Learning Technology

Jen Harvey

*Technological University Dublin, jen.harvey@tudublin.ie*

Martin Oliver

*University College London, martin.oliver@ucl.ac.uk*

Follow this and additional works at: <https://arrow.tudublin.ie/lcart>



Part of the [Education Commons](#)

### Recommended Citation

Oliver, M. & Harvey, J. (2002) What does 'Impact' Mean in the Evaluation of Learning Technology, *Educational Technology & Society*, 5 (3). 18-26, ISSN 1436-4522

This Article is brought to you for free and open access by the Learning, Teaching & Technology Centre at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](#)

# What does 'impact' mean in the evaluation of learning technology?

**Dr. Martin Oliver**

Lecturer, University College London

[martin.oliver@ucl.ac.uk](mailto:martin.oliver@ucl.ac.uk)

**Dr. Jen Harvey**

Distance Education Office

Dublin Institute of Technology

## ABSTRACT

Whilst many projects in Higher Education are expected to demonstrate their impact, quite what this requirement means is often left unspecified. This paper draws on the experiences of the EFFECTS project in an attempt to illuminate this issue. The EFFECTS evaluation framework is used to structure this discussion, which explores the complexities associated with identifying impact in terms of student learning, changes in practice for academics, changes within an organisation and national development. Common themes arising from these areas are then identified. Importantly, while practical issues are considered, the purpose of this discussion is not to 'solve' this complexity – instead, its purpose is to recognise it, and to consider the implications for evaluators of working in such settings.

## Keywords

Evaluation, impact, learning technology, learning, staff development, organisational change

## Introduction

Many projects in Higher Education are expected to demonstrate their impact through evaluation. However, education is a complex process, and the introduction of technology only serves to exacerbate this situation. Consequently, it can be difficult to know where to look for impact, or to recognise it when it happens. In this paper we will explore some of these complexities, illustrating some of the ways in which taken-for-granted assumptions about evaluating projects are, in fact, highly problematic. The discussion will draw on the experiences of the Effective Framework for Embedding C&IT through Targeted Support (EFFECTS) project in order to highlight some of the issues associated with this area, and also to illustrate some of the different forms of impact that can arise in Higher Education. Finally, we will suggest some ways in which these complexities might begin to be addressed, and the kinds of changes that this might entail for the role of the evaluator.

## Background

The EFFECTS project focused on the use of C&IT by academics in Higher Education. It was founded on the premise that targeted support and development for staff would encourage greater use of C&IT, but it also aspired towards appropriate, scholarly adoption rather than introduction at all costs (Harvey & Oliver, 2001). The project involved establishing programmes of staff development at a series of institutions, each of which shared a common set of learning outcomes but was implemented to address specific local issues and needs. Evaluation was an important element of the project from the outset (Oliver *et al.*, 1998), and involved gathering evidence of the project's impact at a number of levels. However, as the project developed, it became apparent that this process was far more complicated than the original evaluation model had suggested. In addition to practical issues, such as the number of case studies produced by participants on the staff development courses, a number of problems arose that brought into question both how and why impact was being investigated. A selection of these problems is presented in the following sections.

## Levels of impact

The EFFECTS project was required to demonstrate impact on several different levels: at the level of individual academics' practice, at an institutional level, and at a national level. Each of these will be considered in turn, below. In addition, a further level will be considered which, whilst of interest to many projects of this type, was not directly achievable within the scope of EFFECTS: the impact on students. This area will be considered first.

## Impact on students

Although not explicitly addressed in EFFECTS, most projects concerned with the introduction of new technology in Higher Education aim to have some kind of impact on students. Often, this is framed in terms of an impact on learning, although perhaps more commonly observed are changes in behaviour (e.g. use of resources), skills (e.g. technical competence) or attitudes (cf. IHEP, 1999). One reason for this is that learning remains an elusive concept, and one that is difficult to measure. The notion of 'learning' is contested, with different theories positioning it as being internal or social, visible and behavioural or internal and mental (c.f., for example, Skinner, 1950 and Kirshner & Whitson, 1997), and so on. It is, thus, impossible to measure the impact of initiatives on learning without taking a theoretical stance on what learning is. Many studies take such a stance implicitly, often simply by concentrating on indicators such as test scores, perhaps not even recognising that this represents a theoretical position.

Whilst many studies use performance on assessment as a proxy for learning, this remains problematic for several reasons. Perhaps most importantly, it is assumed that what has been learnt can be performed; that there is a direct correlation between learning and assessment. This is evidently not the case.

As indicated by Vygotsky's Zone of Proximal Development (1962), children may be able to demonstrate more of their learning when working with a more able peer than working alone; the same effects have been demonstrated for students in Higher Education in a number of contexts (see, e.g., Crook, 1991). This alone suggests that a simple reading of the relationship between assessment and learning is inappropriate. In addition, it is important to recognise that there is much at stake in formal assessment. In order to do well, students will present the best image of themselves as measured by the assessment scheme; they will not present an unbiased, comprehensive illustration of their strengths and weaknesses, because it is not in their best interests to do so. Instead, they will emphasise their strengths and avoid showing weaknesses wherever possible. Moreover, students may find themselves unable to demonstrate that they have learnt something under particular assessment conditions – they may not understand the instructions, feel too pressured, or simply have a 'bad day'. Equally, markers may read more into students' responses than is actually there, and marks will be awarded for correct guesses just as they would for the successful demonstration of learning. Thus assessment is neither comprehensive, unbiased, fair in terms of opportunity or even easy to interpret as a proxy for learning.

The change in educational context and the act of investigating that change can also influence performance on assessment. It has been argued that introducing technology (or any innovation in teaching) fundamentally changes what is learnt, and that as a consequence it is inappropriate to carry out comparative evaluations such as 'before' and 'after' studies of changed courses. As has been pointed out, however, such studies can be and are carried out – however, the complication arises from the fact that such comparisons rely on the same form of assessment being used, and if the nature of learning has changed, then it will have lost its validity for one of the two versions of the course (Oliver & Conole, 1998). Finally, the act of investigating change skews how participants in the research act – this is most widely recognised in the form of the Hawthorne effect, where it is the fact that people are involved in a piece of research and are thus being observed, and not the change being studied, that causes a change in their behaviour.

An important alternative to measuring such forms of assessment is to concentrate on matters of process, rather than of product. Such a shift goes some way towards addressing the problem of inferring from performance back to impact on students. It should be noted, however, for many courses, the introduction of such forms of assessment would create further changes in the students' experience, which would be impossible to disentangle from the introduction of new technology. Nonetheless, where process-oriented assessment strategies are in place, they provide an important new perspective for evaluation. Rather than the changes in process being a problem for the study, they become the focus of the study. The use of learning contracts, self-assessments leading to personal action plans, documented peer-assessments and so on provides a rich set of longitudinal data, often including a number of perspectives and thus supporting triangulation. In many courses adopting such approaches, these forms of evidence are collated and presented as part of a portfolio. This is both a blessing and a curse. On the one hand, large amounts of evidence that would otherwise be unavailable to the evaluator are presented in a format which (with appropriate informed consent) will allow far deeper insights to be developed. On the other, however, the portfolio takes these process-oriented formative assessments and collates them as a summative submission. As noted above, summative assessment is fundamentally about performance, and so the student will have constructed a narrative in their portfolio that shows them in the best possible light. This raises important issues about the reliability of portfolios as a source of data.

There are, of course, other forms of evaluation data that can provide insights into changes in students' learning processes. Direct observation provides one such opportunity. Because evidence gathered in this way are not directly caught up in the formal processes of teaching, learning or assessment, they may prove to be extremely valuable. Mochrie (1998), for example, initially concluded that the provision of randomly generated problems led to improved learning through the promotion of higher-order problem solving, but instead discovered through observation that students were printing out examples and working back from solutions in order to provide practice and a more thorough understanding of typical solutions. Similarly, Jones (1998) identified several instances where evidence that might typically be used for the assessment of processes, such as transcripts of online discussions, was incomplete or even misleading. This was only discovered through an ethnographic-style investigation, which involved observation of and participation in the course. However, ethnographic studies of this type are time consuming, and thus may be impractical in many situations. Similarly, participant-observer data, such as observations by course tutors, will inevitably concentrate on formal interactions with students, omitting the valuable but less visible informal elements of students' experiences.

Changes in skills are arguably at least as difficult to demonstrate as improvements in understanding. Any attempt to measure skills suffers from the same as those identified for assessment, above. Additionally, the whole topic of skills remains conceptually contested, being even less clear than the idea of 'learning'; moreover, failure to use a skill in any given circumstance can just as easily be attributed as to disposition as to the possession or lack of a particular competence (Barnett, 1994).

Perceptions and attitudes are often treated as being relatively easy to investigate; in most cases, it is considered adequate simply to ask participants, for example through an interview or survey. Such changes tend to be viewed as a relatively weak form of impact, however. In part, this stems from the fact that such attitudes are likely to be inconsistent and context dependent, as Social Psychologists have demonstrated (Potter & Wetherell, 1987). By demonstrating this point, Potter & Wetherell have gone on to make a fundamental change to the way in which attitudes are viewed arguing that expressed attitudes are not things that people *have*, they are things that people *do*. Thus the expression of a particular attitude cannot be interpreted as a simple reflection of some inner state, but must instead be understood as a method by which people present a particular image of themselves. Such expressed opinions are intimately tied up with the participant's sense of identity. Whilst gathering attitudinal data is relatively easy, therefore, making sense of it is not, and can require careful exploration in order to appreciate why the participant may be expressing a particular view. (This point is clearly made in a study of practitioners' perspectives of networked learning, as reported in Jones *et al*, 2000. Here a re-interviewed participant clearly acknowledges that their reported account differed from interview to interview.) The implication of this is that the kinds of evidence traditionally used to investigate the impact of a project on students run the risk of being superficial; they do not provide the kinds of privileged insight required to truly demonstrate the extent or implication of any changes that have occurred.

### **Impact on academics**

Many of the issues that arose when considering the impact of projects on students also apply to staff. However, it is rare that staff are involved in learning through an assessed programme (although several EFFECTS courses fell into this category). Consequently, areas such as skills and attitudes are much more commonly investigated. Within EFFECTS, surveys were used to elicit academics' perceptions of the impact that the project had on them. Where interesting issues arose, the evaluation investigated these more fully through interviews with participants. However, it is worth noting that both of these methods were beset by an intractable issue – one attendant on all empirical studies involving the self-report of staff. Much of professional practice remains tacit (McMahon, 2000), and so lies beyond the scope of self-report. However, the fact that things are not noticed, or are not articulated, clearly does not mean that they remain unchanged. As an inevitable consequence, such data will never be able to draw attention to the full range of changes that have occurred.

One important difference between the evaluation of impact on students and on staff is a pragmatic one. The relatively small number of staff means that greater attention can be paid to more subtle, but perhaps more important, indications of impact, some of which may help to alleviate the problem of tacit practice. As noted above, one valuable approach which may not be viable when considering all the students in an institution is observation. This can provide first-hand evidence of changes to practice, some of which may be relatively simple to identify. With technology-related developments, such as the implementation of C&IT within the context of EFFECTS, for example, there is likely to be a tangible difference in the type of resources used. However, observations cannot of themselves explain why the academic changed their practice; nor can they indicate how

prevalent such a change might be. Reasons for the change will be considered in a moment; first, however, we will investigate the problem establishing of prevalence. Two main issues arise here.

The first issue is theoretical: in order to claim that patterns of change have occurred, it is necessary to argue that the sample of cases investigated is representative of the wider population. Representativeness can only be established once the total population has been identified and described. Even then, the choice of features used to select the cases – the characteristics considered to be important – will reflect the bias and assumptions of the evaluator. Thus claims that the findings of studies represent widespread impact are likely to be flawed on at least two counts. Firstly, there is no clear description of the ‘population of academics’ to which the results are supposed to generalise – the claim could be that they are representative of all academics in their department, their institution, their country, or internationally; clearly, ‘typical’ characteristics will not be stable across all of these groupings. Secondly, the characteristics used to describe the population necessarily form an ideology, and as such should be open to critique. Whilst simple characteristics such as age and sex (although not necessarily the more complex concept of gender – Hughes, 2001) are often incorporated, the participant’s social class and economic status, academic history, beliefs and attitudes typically are not. Indeed, it would simply be impractical to account for all of the potentially important characteristics; the influences that shape the way people act are incredibly complex and draw from their complete personal histories, often in ways that remain tacit and unarticulated (Bourdieu, 1977). However, the implication of this is that all claims made about changes to practice are, of necessity, partial and biased, and should be recognised as such.

The second issue relating to prevalence is practical: given that observation became an option because of the small number of staff involved, it may well be counter-productive to attempt to use it on a wider scale. It may, however, be possible to use secondary evidence to supplement direct observation. Within EFFECTS, for example, some evidence for wider uptake of resources was provided by the central units who maintain and support the software. This revealed patterns of use, and of interactions with support staff, which could be used to anticipate sample instances (say, of typical or exceptional behaviour) for observation.

A common type of impact of new initiative in Higher Education is on the role of the academic (e.g. Henkel, 2000), and their expressed attitudes towards (i.e. their use of discourses about) teaching and learning. As with students, evidence of change in such characteristics is hard to gather. Again, however, the small number of staff suggests that some evidence may be obtained from direct observation; other indicators may arise from documentary analysis of outputs such as research reports, course documentation, or discourse analysis of discussions of teaching and learning issues. The adoption of new discourses – such as the introduction of talk about the academic as a facilitator of learning, or use of phrases such as “from the sage on the stage to the guide on the side” – may indicate that their understanding of their role has altered. At the least, it denotes that the view of themselves that they wish to present to others has altered, which in itself should be recognised as a significant development.

Other sources of evidence may provide more direct evidence of a change. For example, a change in their perception of themselves in relation to their discipline may arise from the development of a new area of research interest, evidenced (for example) by a particular conference or journal submission. Where these can be identified, they are likely to be valuable as a source of evidence; however, the problem here lies in anticipating what such changes might be, in order to know where to look for evidence.

### **Institutional impact**

The introduction of new programmes in EFFECTS gave rise to widespread changes within organisations. In part, this was a consequence of the fact that the topics covered in these programmes were closely linked to a wide range of university activities, including those normally associated with librarians, technical support staff, managers and administrators, as well as academics. Such breadth is typical of learning technology developments (McNaught & Kennedy, 2000). Once again, this leads to a tension between breadth and depth in terms of the data collection techniques adopted in the hunt for impact.

As with students and staff, distinct differences can be drawn between different types of evidence in terms of its credibility in demonstrating impact. Within EFFECTS, interviews provided the main body of evidence. Due to the complexity of the work, much the early evaluation simply consisted of documenting which areas staff perceived as having been affected by the work. No systematic attempt was made to gauge the *extent* of this impact; similarly, the credibility of many of these perceptions was frequently taken at face value. The reasons for this were pragmatic: as discussed in previous sections, whilst some simple measures could be adopted here, such

as levels of staff involvement, resource issues, etc., their usefulness as the basis for making judgements is questionable. Numbers of attendees at an event, for example, gives no indication of how the experience influenced those who attended, nor does it provide any insight into the way these individuals would subsequently influence the practices of others in their institution. Thus it was questionable how much insight these additional sources of data would provide, as opposed to simply overwhelming readers with a barrage of measurements.

These issues are echoed in current thinking about the idea of 'learning' in 'learning organisations', which, it has been argued, can never be directly evidenced (irrespective of the volume of data gathered), but only inferred; the claim that an organisation has 'learnt' is simply a narrative constructed in order to present a particular image of the institution (Stewart, 2001). Stewart goes on to argue that the identification of an 'organisation' as distinct from the individuals who compose it is also merely a metaphor. This highlights the importance of considering impact on institutions at two levels: literally, in terms of the changes in relationships between individuals in the organisation, and metaphorically, in terms of the image of the institution that is portrayed.

With these complexities better understood, the later phases of the EFFECTS evaluation made some attempt to question the reliability of the interviews, and to find evidence of the extent of changes. In some cases it was possible to triangulate individuals' claims with those of other groups, or to substantiate them through reference to some other form of evidence. However, it was only as a working model of the impact of EFFECTS on individuals' roles within the institution was developed that it became possible to anticipate which sources of data it would be most valuable to consult.

Within these constraints, there was evidence from the surveys and interviews of changes in roles arising from EFFECTS. Whilst it would have been possible to underpin these claims with evidence of, for example, software licences granted or numbers of support requests made, it was agreed with the project management that the time that this would have taken outweighed the importance of the knowledge that would be gained. These changes in roles included greater propensity for academics to ask for support for the use of learning technologies, an apparent tendency for some academics to view the EFFECTS course teams as their first port of call for all questions, and perhaps most importantly, the development of closer working relationships between different support services in the institutions. This last was largely as a result of the fact that the problems encountered by academics attempting to use learning technology were relevant to several distinct groups, as the following quotes from course tutors show:

"The issues about implementing Perception across the institution has brought together a group of people from registry, management information services, faculty administrators, lecturers, computing services, network services, application teams, you know it has stretched across every single area, it's hard to find an area that hasn't been impacted really."

"God this really is working, there are techies and academics and staff developers all talking together here about the implementation of a managed learning environment."

In spite of the complexities described above, evidence was gathered that suggests that individuals brought about changes to the institutional identity through contributions at an institutional level. This consisted of documentary evidence, together with interview comments describing the role of the individuals in shaping this. Several staff who had been involved in EFFECTS went on to contribute to (and, they claimed, strongly influence) institutional policy. In the interviews, they attributed both their inclination to get involved in these decision-making processes, as well as the expertise that they drew on to influence the discussions, to their prior participation in EFFECTS work. A related issue was that the staff responsible for EFFECTS believed that they had gained credibility internally; invitations to such individuals to participate in policy formation were thus taken as a positive sign of institutional impact.

## **National impact**

The national context, being more complex still, is perhaps the hardest to analyse for impact (Anderson *et al*, 2000). In some cases, there may be an obvious output that is designed to operate on a national level, such as the national accreditation of EFFECTS programmes. These can then be considered to see whether or not they have been successfully established, and monitored to investigate their uptake.

However, this aside, the evaluation of national impact within EFFECTS focused on awareness. This does not appear to be particularly unusual for government-funded projects in Higher Education. In many cases, outputs

such as publications are taken as proxy for awareness, although in practice these provide little or no indication of how widely or influentially they are influencing their audience. Moreover, publications may take years to see print (making measurement through citation impractical for most funded projects) and may still be read many years subsequently.

Perhaps more concretely, contact lists are often used to indicate the number of people reached. In the case of EFFECTS, these were compiled from spontaneous expressions of interest, and from participants in events (such as national workshops) or on programmes. However, just as there are problems evaluating impacts on students' learning, there are difficulties inferring any actual impact from such awareness. Indeed, in several cases, respondents to end-of-project surveys denied being familiar with EFFECTS even though the distribution list had been drawn from participants in workshops and individuals who had directly requested information about the project. Clearly, even when awareness is of interest, its transitory nature means that it must be assessed with caution.

## Discussion

The preceding sections represent an attempt to analyse the concept of impact, as it might apply in the context of evaluating a learning technology project, in a systematic way. What this has highlighted is how problematic such attempts are. Rather than attempting to address each of these issues in turn – which might result in a patchwork of ad-hoc solutions to specific problems – this section will attempt to draw out some of the fundamental issues that underlie the preceding analysis.

Perhaps unsurprisingly, the issues of reliability and validity lie at the heart of many of the problems identified above. The first of these problems is particularly visible in the concerns arising from self-report through survey or interview. However, it also arises for observational data, both in terms of the interpretation of what is observed and the extent to which the samples of observation are representative.

These problems are intractable. The first example illustrates the 'linguistic turn' that has prompted disciplines including philosophy and psychology to reconsider the claims they make and the methods with which they work (e.g. Potter & Wetherell, 1987). It must be recognised that public actions include an element of presentation, so that interviews, policy documents, course materials and so on all form part of a wider attempt to create and maintain a particular image. Whether consciously or not, such data are created in order to convince an audience of something – their bias is inherent; thus they cannot provide privileged access for the evaluator to some absolute truth for any given situation. Where these issues are considered important it may become appropriate to adopt the new empirical methods that have been developed in response to this new awareness of the use of language.

The second example highlights the fact that the objects of study are under-theorised. Since we do not fully understand learning, we cannot provide unambiguous descriptions of when it takes place. Without such a description, the issue of 'representativeness' becomes relatively meaningless. Similarly, what is important in terms of the activities of individuals in a particular organisation is often relatively mysterious and, as with EFFECTS, it is only by undertaking the study that a model can begin to be built – meaning that the issue of sampling can only begin to be considered once data has already been collected. The initial EFFECTS model, which simply elaborated the various levels used to structure the study, did precede formal data collection. However, as has been shown, the reality of the situation – particularly in terms of things like institutional culture – was more complex, and the levels interacted with each other. Thus the initial hypothesis only gained credibility as a model through the data collection process and the subsequent refinement of our understanding of it. Recognition of this lack of adequate descriptions is a necessary first step; attempts to develop models through evaluation may help further, but these will only serve to explain, not solve, the problem.

A further problem runs through all the issues described in this paper: the inference of causality from evidence. As highlighted earlier in this discussion, sensitivity to the social function of discussions and documentation means that care must be taken when working with interviews and so forth. One consequence of this is that we can claim little more than that "interviewees stated...", or "the documentation implies...". Taken together with the lack of theories and models, and the complexity of the contexts under study, the warrant with which conclusions can be drawn becomes increasingly unsteady.

It is well recognised that qualitative and quantitative analyses should be judged differently, with issues of credibility, thoroughness of description, consistency, relevance to its audience and reflexivity determining the

quality of qualitative studies (Mays & Pope, 2000). Interestingly, such criteria resemble the principles underlying Utilization-focused evaluation (Patton, 1997). This particular approach recognises that evaluation is a social phenomenon, and proposes that studies be judged on the basis of their fitness for purpose – did they help the sponsors of the study do what they needed to do? However, it is important to recognise that using such a question as the basis for judging quality goes further than the shift from quantitative to qualitative criteria. Inherently, it recognises that the criteria – whatever they may be – will be interpreted differently by each audience depending on their local beliefs and context (Borum & Hansen, 2000). It could be argued that this represents an attempt to bring the well-recognised epistemic struggle between positivist and relativist approaches to a more post-modern position (Oliver, 2000).

Superficially, such a move might seem to imply the abandonment of standards, or even a slide towards a situation in which the sponsors of a study set not only the agenda but also determine what the results mean. However, such pessimistic readings are far from being the only conclusions that can be drawn about this development. Instead, it is possible to see this more as having a series of implications. What it implies is that evaluators must:

- pay greater attention to the beliefs of the various groups involved in the study, and the way that these define the situations they describe.
- be more sensitive to the goals of the various groups, and the way that these influence the data that are created and presented.
- attend to the potential educative role of their work, which represents a valuable opportunity for dialogue (cf. Freire, 1970).
- recognise that their own reporting is a form of narrative, based on personal understanding (perhaps negotiated amongst a small team), which serves to present themselves as credible and professional evaluators as well as helping to inform an audience.

Needless to say, the beliefs and goals of the evaluator are one of those that must be represented and negotiated within the first two of these points.

These complexities are well illustrated by the topic of ‘impact’. Use of this term by the sponsors reflects a rational, causal model of interventions that may be markedly different from the complex, messy world of appropriated tools and re-interpreted policies experienced by the academics. The sponsors wish to find evidence of impact in order to justify their investment, whereas the academics may simply have wished to use the funding to improve the students’ experience. Discussion of the experiences of students may lead different groups to appreciate their own role in the educative process, and to rally around the ideal of helping students to make the most of their studies. Additionally, the evaluator’s report will either play to the sponsors by claiming that there is such a thing as impact and that they have found it, or that ‘impact’ is an over-simplistic concept and that they have gone beyond it. Either way, it will typically summarise their synopsis of other people’s experiences, and will only acknowledge shortcomings as a ‘limitations’ section designed to enhance their credibility as a reflective practitioner.

## Conclusions

The brief for the evaluation of EFFECTS concerned the idea of measuring impact, reflecting the priorities of the national programme and the importance attached to accounting for funding. However, what ‘impact’ meant, and how it would be identified, remained unaddressed. This paper has attempted to explore this issue, and in doing so, has highlighted a number of problems both with the concept and its evaluation.

The complexities of ‘impact’, as played out in the specific context of supporting academics’ use of learning technology, are numerous. A project such as EFFECTS touches upon a range of areas, including students’ learning, universities’ organisational behaviour and academic identity construction, all of which are poorly understood. The interplay between these simply adds to the complexity of the topic of study. In such a situation, it is hard to identify what might count as having changed, let alone to attribute a cause to the phenomenon. As a consequence, there is a need for evaluators to be more explicit about their understanding of the focus of their study. Unless they explicitly start to build models – and theories – of the phenomena they are investigating, there will be no credible warrant for the conclusions they drawn. Given our growing appreciation of the complexity of the topics under investigation, calls to common sense understanding no longer provide convincing justifications for claims.

Moreover, throughout the discussion, the social nature of evaluation has been highlighted. The investigation cannot be perceived as a value-free enquiry into truth; as recent theories of evaluation have stressed, it is an inherently political activity (e.g. Patton, 1997). What has been argued here is that this should lead to a deeper sensitivity towards the rhetorical nature of both the evidence that evaluators encounter and the reports that they produce. With an increased awareness of such matters, and of the potential for educative dialogues with and between the different groups participating in the study, ethical issues become a priority. The varied agendas of the evaluation sponsors, the study participants and the evaluator must be engaged with and explored, and the evaluator must take up a thought-through position that allows them to present these in good conscience to the people who are to act on the study.

Finally, the notion of impact itself has been elaborated, both through the experiences of the EFFECTS project and in terms of the issues facing evaluators. The attempt to evaluate 'impact' serves a legitimating end for the sponsors, who wish to justify their use of funds and claim credit for the benefits that their programme has 'caused'. In such circumstances, quite what 'impact' is remains usefully ambiguous; this allows those being studied to present themselves in the best light possible (so as to maintain their own sense of worth) by finding any data that can be construed as evidence of improvement. All that then remains is to construct credible rhetoric to justify the links between these changes and the funded programme.

This interpretation is deliberately cynical. Such an outcome is not, however, inevitable. Indeed, part of the point of this discussion, and potentially the most important conclusion to draw from this paper, is that there is an alternative for evaluators that will enable them to carry out their work thoughtfully and sensitively. By explicit theorising and through socially-aware dialogue, it remains possible for evaluators to construct alternative accounts, ones which may also pass the tests of credibility and usefulness, but which represent the values and beliefs of the various groups involved rather than simply conforming to any one dominant agenda.

## Acknowledgements

We would like to thank the anonymous peer reviewer whose critique of the initial outline of this paper helped improve and expand on it considerably.

## References

- Anderson, C., Day, K., Haywood, J., Land, R., & Macleod, H. (2000). Mapping the Territory: issues in evaluating large-scale learning technology initiatives. *Educational Technology & Society*, 3 (4), [http://ifets.massey.ac.nz/periodical/vol\\_4\\_2000/anderson.html](http://ifets.massey.ac.nz/periodical/vol_4_2000/anderson.html)
- Barnett, R. (1994). *The Limits of Competence: Knowledge, Higher Education and Society*, Buckinghamshire: OU Press.
- Borum, F., & Hansen, H. (2000). The local construction and enactment of standards for research evaluation: the case of the Copenhagen Business School. *Evaluation*, 6 (3), 281-299.
- Bourdieu, P. (1977). *Outline of a Theory of Practice* (2001 edition; translated by R. Nice), Cambridge: Cambridge University Press.
- Crook, C. (1991). The Zone of Proximal Development: Implications for Evaluation. *Computers and Education*, 17 (1), 81-91.
- Freire, P. (1970). *Pedagogy of the Oppressed* (20<sup>th</sup> Anniversary Edition, 2000), New York: Continuum Publishing.
- Harvey, J., & Oliver, M. (2001). EFFECTS external evaluation, *Unpublished project document*.
- Henkel, M. (2000). *Academic Identities and Policy Change in Higher Education*, London: Jessica Kingsley.
- Hughes, G. (2001). Exploring the Availability of Student Scientist Identities within Curriculum Discourse: an anti-essentialist approach to gender-inclusive science. *Gender and Education*, 13 (3), 275-290.

- IHEP (1999). The Institute for Higher Education Policy. *What's the Difference? A review of contemporary research on the effectiveness of Distance Learning in Higher Education*. Institutional report, <http://www.ihep.com>.
- Jones, C. (1998). Evaluating a collaborative online learning environment. *Active Learning*, 9, 31-35.
- Jones, C., Asensio, M., & Goodyear, P. (2000). Networked learning in higher education: practitioners' perspectives. *ALT-J*, 8 (2), 18-28.
- Kirschner, D., & Whitson, J. (1997). *Situated Cognition: Social, Semiotic and Psychological Perspectives*, New Jersey: Lawrence Erlbaum.
- Mays, N., & Pope, C. (2000). Quality in Qualitative Health Research. In Mays, N. & Pope, C. (Eds.) *Qualitative Research in Health Care*, London: MBJ Publishing, 89-102.
- McMahon, A. (2000). The development of professional intuition. In Atkinson, T. & Claxton, G. (Eds.) *The Intuitive Practitioner: On the Value of Not Always Knowing What One is Doing*, Buckingham: Open University Press, 137-148.
- McNaught, C., & Kennedy, P. (2000). Staff development at RMIT: bottom-up work serviced by top-down investment and policy. *ALT-J*, 8 (1), 4-18.
- Mochrie, R. (1998). A Diet of Carrots: Autonomy in Learning Mathematics for Economics. In Moge, N. (Ed.) *Evaluation Studies*, Edinburgh: Learning Technology Dissemination Initiative, <http://www.icbl.hw.ac.uk/lti/evalstudies/escarrots.pdf>
- Oliver, M. (2000). An introduction to the evaluation of learning technology. *Educational Technology & Society*, 3 (4), 20-30.
- Oliver, M., & Conole, G. (1998). The evaluation of Learning Technology - an overview. In Oliver, M. (Ed.) *Innovation in the Evaluation of Learning Technology*, London: University of North London, 5-22.
- Oliver, M., Conole, G., Phelps, J., Maier, P., Wilkinson, D., & Bailey, P. (1998). *The EFFECTS Evaluation Framework: a transferable model for collaborative projects*, EFFECTS report no. 1, University of North London.
- Patton, M. (1997). *Utilization-focused evaluation*, London: Sage.
- Potter, J., & Wetherell, M. (1987). *Discourse and Social Psychology: Beyond Attitudes and Behaviour*, London: Sage.
- Skinner, B. (1950). Are theories of learning necessary? *Psychological Review*, 57, 193-216.
- Stewart, D. (2001). Reinterpreting the Learning Organisation. *The Learning Organization*, 8 (4), 141-152.
- Vygotsky, L. S. (1962). *Thought and Language*, Cambridge, MA: MIT Press.