

12: Evaluation: A practical guide to methods

Philip Crompton

This article seeks to cover the principles for conducting an evaluation whether it is a small or a large project. An understanding of the theory and background to evaluation is beneficial in order to better plan, design and conduct an evaluation programme. Above all there is no substitute for having a clear purpose for an evaluation, defining the right question is a prerequisite.

Definition

“Evaluation is the collection of, analysis and interpretation of information about any aspect of a programme of education or training as part of a recognised process of judging its effectiveness, its efficiency and any other outcomes it may have.”

Mary Thorpe in “Handbook of Education Technology” (Ellington, Percival and Race, 1988)

As the authors point out, Rowntree (1992) makes comment on this definition as follows:-

1. Evaluation \neq Assessment
2. We should not ignore the unexpected.
3. Evaluation is a planned systematic and open endeavour.

Evaluation and assessment although often used interchangeably, refer to different levels of investigation. Evaluation is concerned at the macro or holistic level of the learning event, taking into account the context of learning and all the factors that go with it, whereas assessment can be seen as the measurement of student learning and is one of the elements that go into an evaluation, the micro-level. One aspect of any sound evaluation is the allowance for the unexpected. Above all an evaluation is a designed and purposeful enquiry which is open to comment.

Approaches

There are initially two distinct approaches to evaluation

- (i) the agricultural/botanical or scientific approach or
- (ii) the social/anthropological or illuminative approach.

The scientific approach is concerned with the measurement of the effects of specific variables against the resultant outcomes. It seeks to examine the achievement of stated goals in relation to a learner’s pre-knowledge and skills. The approach is geared towards the measurement of the efficiency of the educational intervention as well as the effectiveness of the learning outcomes.

The illuminative approach on the other hand seeks to examine and explore the process of the educational intervention. The techniques, as such, are therefore more qualitative, some might say, more subjective in nature as they call on personal judgements.

Though these approaches appear to be at either end of a continuum, it is possible to make use of both within the complexity of what is educational research. A selected mid-point will be governed by that which is being evaluated as well as the reason behind the evaluation. It is not uncommon to see a mix of techniques from both approaches combined in an evaluation project. The degree of combination will depend largely on a process of negotiation between the evaluator and the instigator of the evaluation as well as the environment and the time frame in which the evaluation is being performed.

Whichever approach or combination of approaches is used, the evaluation should always be a clear well thought out undertaking. The more effort that goes into the pre-planning of a piece of evaluation the better the evaluation. Before conducting any evaluation it is important to not only have defined that which you are trying to investigate but also how you are going to go about it.

Levels of evaluation

Romiszowski (1988) differentiates between the scope (Levels of Evaluation) and depth (Levels of Analysis) of an evaluation.

LEVEL I	-	Project Level
LEVEL II	-	Curriculum Level
LEVEL III	-	Unit Level
LEVEL IV	-	Learning Step Level

It is necessary to predetermine at which level the evaluation is to be conducted. Evaluation is of great importance in all aspects and stages of teaching and learning. All too often it is seen as the last activity and yet as Laurillard (1993) states it is an iterative process and should take place at every stage in the design, production, implementation/integration of a new educational intervention (e.g. using LT materials) whether it be as a complete course, part of a course, or a particular session or teaching aid.

There are two possible stages for evaluation;

- a) the evaluation of the IT intervention in isolation, and
- b) the evaluation of the IT intervention within the course itself.

The evaluation of a piece of courseware in isolation will tend to focus inwardly on various aspects of the software itself. It will look at aspects like navigation, screen design, text layout, etc. It will examine the scope of the coverage of the material content, e.g. is the subject presented to such a breadth and depth that its use is not limited to beginners?

The evaluation of the courseware within the course itself will allow us to examine other factors which will lend themselves to the successful integration of the product within the course. These will be:

- educational setting
- aims and objectives of the course
- teaching approach
- learning strategies
- assessment methods
- implementation strategy

The iterative nature of evaluation should then assist in making the learning experience both more efficient as well as more effective as the feedback is used to continuously improve matters.

It is important to recognise at the outset that evaluation is a time consuming exercise and better planning can save time and effort in the actual evaluation process. It is more advisable to do a small evaluation well than to try to do a large evaluation and run out of time. Remember that an adequate amount of time needs to be allocated to both the analysis and writing up stages.

Effectiveness, efficiency and relevance

Two main concerns for evaluation in the context of the TLTP initiative have been the need to justify technology-based learning on the grounds of effectiveness and efficiency. A third factor should be added to the list and that is relevance; the applicability and appropriateness to the intended employers and users of the technology, the teachers and students, even the department or institution. Whereas Romiszowski (1988) includes this under factors of effectiveness, there is a case for treating it separately. Project Varsetile (Allen, Booth, Crompton & Timms, 1996) encountered technology that would score highly on both grounds of effectiveness and efficiency and yet at the same time be wholly inapplicable for the given context of instruction and learning. Successfully integrating or embedding the CBL package within a course is dependant on relevance.

The following are factors open to consideration and investigation by an evaluation study.

Factors in achieving Process (EFFICIENCY)

- Cost
- Structure
- Alternatives
- Resources
- Organisational

Factors in achieving Objectives (EFFECTIVENESS)

- Cognitive and/or Physical Skills
- Affective

Factors in achieving Satisfaction (RELEVANCE)

- Affective
- Scope of the Content
- Degree of Effectiveness And Efficiency

The evaluation of these factors will lend themselves to improvements in both the course, in terms of aims, objectives and content, and the procedures for the design and development of the course itself.

Techniques

This section will look at the various instruments available to the evaluator and while the list is not exhaustive it is representative of the main techniques used in any evaluation.

Questionnaires

This is probably one of the easiest and simplest instruments to create for data collection. However this apparent simplicity belies the fact that creating and organising your questions can be quite arduous and time consuming. You should be clear about the areas and concerns that the questionnaire is to cover. Do not attempt to ask everything as this will tend to make the questionnaire too long and by so doing decrease the willingness of the respondent to answer fully and thoughtfully.

Designing a Questionnaire

In designing the layout of the questionnaire, leave enough space for responses to open-ended questions but at the same time keep such questions to a minimum if possible. These areas are often better left to semi-structured interviews where a fuller answer can be sought. If these are necessary, keep them to short-answer questions allowing the respondent some space to elaborate on their response.

- Using the standard Likert five point scale makes it easier to collate the data. Many an evaluation can be hampered by the sheer volume of data to be transcribed and input for analysis.
- Keep the questionnaires anonymous as this will enable a freer response than if the students can be identified, in which case they might not be as forthcoming.
- As questionnaires are less time consuming to administer than other instruments of evaluations try not to over use them. Students can become uncooperative if every time they do something they are presented with a questionnaire.
- Take into account the advantages and disadvantages of using questionnaires.
- **Above all pilot the questionnaire before using it on your target group.**

Advantages

- Covers a lot of information in a simple format.
- Allows for anonymity, perhaps eliciting more genuine responses.
- Fairly simple to administer.
- Easy to collate and analyse.

Disadvantages

- Difficulty in getting a useful response rate (not everybody returns the questionnaire).
- Open ended questions are more difficult to categorise.
- Responses may be based on the more memorable events (either good or bad).
- Any follow-up must be done through interviews.
- Necessary to keep the questions focused and to ask the right kind.
- Can become merely “feel good” tick sheets.

Semi-structured Interviews

This evaluation is best used after you have collated and analysed your questionnaire. This will then allow you to obtain an elaboration on important points arising out of your questionnaire. It even allows you the opportunity to cover those points you did not manage to cover in the questionnaire.

Developing your interview questions

While developing your questions from points arising out of your questionnaire do not feel confined by these, allow the interviewees to raise points of concern to them about the learning event. By so doing you may discover points that you yourself had not thought of. However, do not allow yourself to be side-tracked too much.

- Do not forget that interviews tend to collect a lot of data by sticking to your prepared and scripted questions and keeping your interviewee(s) focused you will resist collecting too much unwanted data.
- Interviews may be conducted individually or in groups. One of the benefits of group interviews is that one comment may encourage others.

Advantages

- Allows for elaboration of answers.
- Potential for revealing points not previously considered.

Disadvantages

- Takes time to arrange the interviews and to conduct them.
- Potential for collecting too much data, some of which may be unwanted.
- Analysing the responses.

Confidence Logs

These were developed by the TILT Project (University of Glasgow) and are extremely useful in gauging a student's confidence on particular points of a given subject area. These are useful as self indicators of learning rather than anything more concrete. Although their relationship to the actual learner is unproven they can be a useful guide to a student's self-appraisal of his or her own knowledge. It is beneficial within a learning event if the learner's confidence is higher in order that they may actually learn. Under confidence as much as over confidence can be an inhibiting factor as a foundation for learning.

Example:

This is taken from an evaluation carried out on a piece of software produced by a TLTP Economics consortium. The Confidence Logs demonstrated an increase in the learners self-assurance with the material. On average their confidence moved from a feeling of “Some-Confidence” to that of “Confidence”. Although this does not necessarily translate into increased attainment scores it is a necessary element to any learning environment.

	Very Confident	Confident	Some Confidence	Little Confidence	No Confidence
Pre	0	4	11	2	2
Post	1	12	4	1	0

Table 1: Confidence Logs

1Crichton (1995), cites Belcha (1992), in her preliminary observations of an earlier version of *WinEcon* states that “the subjective measure used here has the strength of reflecting a student’s own judgement about whether or not the software was helpful in learning course material.”.

Observations

Again, as with any instrument the better prepared and the more focused an observation is the more likely it is that useful data can be collected. Although with video-taped recordings it is possible to make observations at a later date, it is far better to have a sense of what one is seeking to observe whether it be inter-personal communications, skills or behaviour patterns (e.g. study strategies - note taking, etc.)

Use of a structured observation sheet is essential whether the session is video-taped or not. If video-taping is not possible, then the use of a tape-recorder is useful as students verbal interactions and comments may be synchronised with the notes and comments from the observation sheets.

It is useful to have a code or shorthand that can help in recording your observations. This will help you focus on what it is that you are trying to observe while at the same time allowing for those unexpected but enlightening events which may occur. If you are conducting the observation exercise with a colleague make sure that you are both clear on the meaning of the shorthand code.

Example:

This particular observation sheet was created for use in the evaluation of a piece of Japanese CAL software. A pair of students were video taped using the program and their actions were noted on the observation log when the tape was later analysed.

<u>Evaluation Exercise</u>			
Date:		Institution:	
Lecturer:		Evaluator:	
Hardware:		Program:	
Student Name(s):		Time:	
 <u>Observation Log</u>			
CODE for interactions:			
TS	=	Teacher to Student	SS = Student to Student
GD	=	Group Discussion	SC = Student to Computer
GC	=	Group to Computer	SI = Student Individually
#	Time	Code	Comment

Student Profiles

It is useful to start any evaluation by obtaining a profile for each student as a baseline against which any future changes in both attitude and opinion may be referenced. Within a Student Profile questionnaire it is appropriate to ask about the learner’s attitude towards computer-based learning as well as their use of computers in general. At the same time, questions concerning the learner’s academic background in the particular and related areas of

the content subject under evaluation. If for instance in an evaluation programme on CBL in economics, information about the learner's level of knowledge of not only economics (e.g. Standard Grade or Higher, O or A levels etc.) but in the related areas of maths and statistics would be useful.

Profiles are also useful when selecting evaluation groups by the use of stratified random sampling through which it is possible to obtain groups of learners that meet specific criteria. When wanting to make sure that there is a range of student abilities in a certain group the use of a student profile questionnaire can be used as a basis for this selection.

Pre and Post Tests

This is perhaps one of the most difficult techniques in terms of educational research. The idea of performing a pre-test before the students use of a particular program followed by a post-test in order to measure any educational gain is not as simple as it might first appear.

The concern is that of test validity which is fundamental to the purpose of any testing. Cohen and Manion (1980) raise the point that:

“Conditions that threaten to jeopardise the validity of experiments... are of greater consequence to the validity of quasi-experiments (more typical in education research) than to true experiments...”

The problem is one of both ‘internal validity’ and ‘external validity.’

Internal validity is concerned with whether the educational intervention makes any difference to the outcome in question. Whereas external validity is concerned with how far the results are generalisable and in what settings and with what population. There are different factors which affect validity which fall outside the scope of this article. However one of the main problems within higher education is that it is becoming increasingly difficult if not time consuming to create a ‘true’ experimental design.

Notwithstanding the potential pitfalls and accepting the limitations of this technique indicators of learning gains may be obtained, and while not being generalisable may be useful when taken together with other findings.

Example:

This example is once again taken from the evaluation of the economics software package *WinEcon* together with *Economics in Action* and the two recommended coursebooks that these software programs are based on. The students were randomly split into three groups and each group were given a different learning resource for each test. They were tested at both the beginning and at the end of each session.

	TEST 1	+/- %	TEST 2	+/- %	TEST 3	+/- %
Group 1	WinEcon	+11.11	EIA	+10.83	Books	+11.31
Group 2	EIA	+11.81	Books	+10.00	WinEcon	+8.57
Group 3	Books	+25.69	WinEcon	+17.00	EIA	+2.04

The pre- and posts-tests results appear to indicate that the software was certainly no worse than more traditional forms of book-based learning. The first test appeared to indicate that the possible automatic note taking by the students seemed to have had a beneficial impact which is somewhat equalised as all the students adopt this approach in the subsequent two tests when using either piece of software.

A one-way analysis of variance was carried out on the data in order to identify whether any group was better at economics. The pre-test scores of each group were subjected to this analysis but there were no significant differences. The nearest to any effect occurred in Test 1, though this was not statistically significant. No matter which media was used the students performed equally well, however, with such a small sample small changes in performance were undetectable.

Checklists

It is useful to document all aspects of the evaluation procedure and check this off once completed. Tessmer and Harris (1992) make great use of checklists as templates for conducting evaluations. Having a checklist of the stages within an evaluation can help in the shaping and structuring of the overall evaluation design, besides their use in the evaluation itself.

The TLTP project TILT have developed a useful checklist of all learning resources available to students. At the end of a course or programme the students complete the checklist indicating what they found useful. It could include such resources as books, handouts, tutor's time, other students, CBL packages etc.. This is useful in indicating those resources which a student uses and finds most helpful over the period of a course.

Another excellent source is the "Telematics Applications for Education and Training USABILITY GUIDE, Version 2" by Smith & Mayes (1995) as it contains references to other evaluation sources. It is freely available over the world wide web via URL "<http://www.icbl.hw.ac.uk/tet/usability-guide/>".

Example:

This checklist can be used when initially planning an evaluation in order to begin to create an evaluation plan. It could then be written up as the basis for a contract, either formal or informal, for any evaluation work to be performed.

INITIAL EVALUATION ASSESSMENT CHECKLIST

- Who?** (Target — know your audience)
Who is the evaluation for?
- What?** (Area — understand what it is you are evaluating)
Process (Efficiency)
Outcome (Effectiveness)
Combination of both (Relevance)
Purpose (Validate, Improve or Condemn)
- When?** (Timing — don't start until you are ready)
have you defined a question?
will the findings have any effect?
benefits outweigh costs
- How?** (Techniques — what is most appropriate)
Questionnaires
Interviews
Confidence Logs
Observations
Student Profiles
Pre-Tests and Post Tests
Inventory Learning Checklists

Summary

The effort that is put into the design of any piece of evaluation will pay rich dividends, however, defining the right question is always the key starting point. There are degrees of correctness of definition but this should always be something that is measurable and possible within the time frame in which you are working. Each evaluation will have its own combination of costs and benefits (see Breakwell and Millward, 1995), and estimating them is all part of the work of an evaluation.

References

- Allen, P, Booth, S, Crompton, P, and Timms, D (1996) "*Integrating Learning with Technology*", University of Stirling, Stirling

- Belcha, B J (1992) Learning Styles, “*Economics, and Software*”. Paper presented at the 1992 American Economics Association Meeting, New Orleans quoted in Crichton, P “*Evaluating WinEcon,*” CHEER, Vol. 9, Issue 1 1995
- Beynon, J, and Mackay, H. (Ed) (1993) “*Computers in Classrooms: More Questions than Answers*”, Falmer Press, London
- Davies, I K, (1971) “*The Management of Learning*”, McGraw-Hill, London
- Draper, S W, Brown, M I, Edgerton, E, Henderson, F P, McAteer, Smith, E D, and Watt, H D (1994) “*Observing and Measuring the Performance of Educational Technology*”, University of Glasgow
- Draper, S, Brown, M J, Henderson, F P, and McAteer, E (1995) “*Integrative evaluation: An emerging role for classroom studies of CAL*”, unpublished
- Ellington, H, Percival, F, and Race, P (1993) “*Handbook of Educational Technology*”, Kogan Page, London
- Rowntree, D (1982) “*Educational Technology in Curriculum Development*”, Athenaeum Press Ltd, Newcastle upon Tyne
- Laurillard, Diana (1993) “*Rethinking University Teaching*”, Routledge, London
- Romiszowski, A.J. (1988) “*The Selection and Use of Instructional Media*”, Kogan Page, London
- Smith, C. and Mayes, T, (1995) “*Telematics Applications for Education and Training USABILITY GUIDE*”, Version 2, DGXIII 3/c, Commission of the European Communities, Brussels
- Tessmer, M and Harris, D (1992) “*Analysing the Instructional Setting*”, Kogan Page, London
- Wade, W, Hodgkinson, K, Smith, A, and Arfield, J (1994) “*Flexible Learning in Higher Education*”, Kogan Page, London

Further reading

- Anderson Lorin W. and Burns, Robert B. (1989) “*Research in Classrooms*”, Pergamon Press , Oxford
- Bennett, Clinton, Foreman-Peck, Lorraine, and Higgins, Chris (1996) “*Researching into Teaching Methods in Colleges and Universities*”, Kogan Page, London
- Breakwell, Glynis and Millward, Lynne (1995) “*Basic Evaluation Methods*”, BPS Books, Leicester
- Cohen, Louis and Manion, Lawrence (1994) “*Research Methods in Education,*” (4th edition) Routledge, London
- Creswell, John W. (1994) “*Qualitative & Quantitative Approaches*”, Sage, London
- Forsyth, Ian, Jolliffe, Alan and Stevens, David (1995) “*Evaluating a Course*”, Kogan Page, London
- Hopkins, David (1992) “*Evaluation for School Development*”, Open University Press, Buckingham
- Tessmer, Martin (1993) “*Planning and Conducting Formative Evaluations*”, Kogan Page, London

The evaluation instruments referred to in this chapter are included in Appendix 2.

Evaluation is also addressed in the following chapters: Chapter 11 - a conceptual introduction; Chapter 3 - the role of evaluation in the overall process of implementation, and Chapter 7 - a practical guide on how to evaluate LT materials that you may be considering using in your teaching.