Evaluating Virtual Learning Environments: what are we measuring?

Mary C Dyson and Silvio Barreto Campello
Department of Typography & Graphic Communication, The University of Reading, UK

M.C.Dyson@rdg.ac.uk
S.B.Campello@rdg.ac.uk

Abstract: A basic framework is proposed to distinguish between the many ways in which Virtual Learning Environments (VLEs) can be evaluated. This includes the purpose of the evaluation, the type of methods that might be used and the measures employed. The framework is not intended to cover all applications but offers one means of structuring a review of past studies or may provide guidance on the type of study to conduct. A pilot study is introduced which compares an online course using different platforms which aims to measure engagement, participation and achievement of goals.

Keywords: purpose, methods, measures, usability, learning.

1. Introduction

Institutional strategies for the development and use of ICT in Higher Education in the UK are now in place, as recommended by the Dearing Report. At the University of Reading, one of the principles of the strategy is that new technologies should encourage rethinking of pedagogical aspects of teaching, learning and assessment. The Virtual Learning Environment Blackboard was purchased in August 2000 and there are currently approximately 100 courses online. The Universidade Federal de Pernambuco (UFPE) in Brazil, the home university of the co-author, has developed their own system, VirtusClass.

Evaluation provides feedback for course developers on teaching and learning and is an important part of quality assurance. However, constraints on time and possibly expertise preclude most developers from detailed studies. Whilst it is still important to carry out evaluations of individual courses, looking for more general principles derived from experiments can provide guidance in the design and development of VLEs. Such research may also address issues that are not covered in many evaluations.

1.1 Outline of paper

Having discovered the range of evaluation studies of VLEs and related learning technologies reported in recent relevant journals, this paper outlines a basic framework to distinguish between evaluations. The framework was developed to provide the context for discussion of a pilot study looking at the effects of group orientation on students' engagement, participation and task engagement. Dimensions are therefore identified that may be relevant to this study. However, the framework is intended to be of more general use. It may offer a means of structuring a review of past studies, for example, to identify the most relevant, or may provide guidance on the type of study to conduct.

The pilot study is introduced by defining the theoretical position underlying the research. The variables chosen for investigation are outlined and the pilot tests the appropriateness of these variables. Outcomes are briefly described with suggestions as to how the design of future studies can be informed by these results.

2. The nature of evaluations

In considering literature on the evaluation of VLEs or similar technologies, it is apparent that there are many different approaches to studies. A useful framework has been devised by Oliver (1997), which provides a comprehensive guide to the evaluation of the use of educational technology. This report is used as a starting point for discussion of the factors that are considered relevant to the current paper.

It is possible that the term 'evaluation' may be restrictive in the current context. Evaluation has been clearly explained by Oliver (2000) as 'the process by which people make value judgements' and when applied to learning technology, he suggests that this is often the educational value of innovations or practical issues in introducing new teaching methods and resources. Whilst the overall objectives of such evaluations are likely to be identifying what may improve learning, some evaluations have specific outcomes, whilst others aim for more general relevance. Oliver (1997) is well

aware of this distinction, which is built into the five purposes for evaluation (described below). A more marked distinction is made in the current paper by suggesting that it may be helpful to regard some studies as 'experiments' and some as 'evaluations'.

2.1 Purpose of evaluation

2.1.1 Roles

The starting point for distinguishing between different evaluations is naturally the purpose of the study. Oliver (1997), based on Draper, Brown, Henderson and McAteer (1996), identified five roles for evaluation: formative, summative, illuminative, integrative evaluations and quality assurance. Quality assurance is undoubtedly a specific purpose for evaluations. However, within the field of Human Computer Interaction (HCI) formative and summative evaluations are characterised by the stage in the development process at which they occur (Preece, Rogers, Sharp, Benyon, Holland and Carey, 1994), although this also defines their purpose. Explanations of illuminative and integrative evaluations illustrate the close relationship between purpose, approach (e.g. experimental versus ethnographic) and measures. For instance, illuminative evaluations are described as being primarily ethnographic, as opposed to experimental. Their purpose is to discover issues considered relevant by participants. Integrative evaluations are closely related to illuminative and aim to provide specific guidance on delivering effective teaching and learning.

2.1.2 Experiments

Four of these five roles are identifying problems, describing and interpreting events, rather than studies, which may test a single well-defined question (summative evaluations) and provide results of more general relevance. These objectives provide criteria for distinguishing between evaluations and experiments. A case study of web-based support for a campus-based course (Holt, Oliver and McAvinia, 2002) departed from the more usual focus on the particular system and cautiously discussed the wider implications of the study. A more obvious example of a study that would qualify as an experiment is Woods and Keeler (2001), which assessed the effect of adding audio to emails. The specific research questions were whether the audio messages would increase the frequency of student participation and length of utterances in online asynchronous group discussion and whether they would also result in more favourable student perceptions.

The classical design of an experiment is a comparison of conditions, sometimes with a control group. This was carried out by Woods and Keeler (2001) in the study referred to above. They compared three levels of audio messaging (weekly, monthly and every other month) with no audio messages (the control group). These designs can be problematic in natural settings due to difficulties in achieving comparable situations, avoiding contact between groups where they may share material specifically intended for one group, and possible ethical problems such as depriving some people of a potentially richer learning environment.

2.1.3 Usability versus learning

Another dimension that separates studies is the approach adopted by the specific discipline. Whilst studies within the educational field aim to assess students' learning outcomes, situating the evaluation within an educational context that incorporates assessment, an alternative objective is to measure usability of the system and its tools, drawing on HCI research. An example of this is Chang (2001) who investigated whether a web-based learning portfolio enhances learning outcomes by measuring the usability of the system.

Definitions of usability vary but there are similarities in the type of variables they tend to measure. These include effectiveness, efficiency and satisfaction (ISO 9241), ease of remembering and error rate (Nielsen, 1993). Commonalities among definitions found in the literature are making the use of a system easier and more comfortable for the users, whilst guaranteeing a high level of productivity.

However to measure the level of productivity in the field of learning technologies may be particularly difficult. The crucial point is the conception of learning that underlies evaluation. Typical measures used to evaluate the usability of a system, response time, accomplishment of tasks, error rate, etc. are suitable for a large range of systems and even for Computer Assisted Instruction Systems (CAIS) or Intelligent Tutor Systems (ITS). However, if learning is conceived as a matter of process, during which a transformation of knowledge occurs, such measures say nothing about how new knowledge has developed and what is necessary to support this development.

As all activity within a VLE is carried out through the interface, it is important to examine

how this may support learning. However, it is unhelpful to take the evaluation out of the learning context to focus only on ease of use of the system. The purpose of the evaluation should determine what is measured but it is the conception of the investigated phenomena that defines what is actually observed. In usability research the focus of the studies seems to be the individual using the system. Cultural factors that surround the use of the system are not included in the analysis. The context is merely a scenario that provides information about the task performed but is not part of the experience. Usability and learning may be combined in a single study, but each will have their own individual measures. How measurement is conducted is affected not only by the specific variables, but also by the circumstances surrounding the evaluation.

2.2 Methods

2.2.1 Interpreting results

Employing experimental methods to evaluate learning technologies is often considered inappropriate due to the difficulty of controlling variables that may affect outcomes (reviewed in Jones, Barnard, Calder, Scanlon and Thompson, 2000). However, in a natural context, where the technology may be only one part of a course, other evaluation methods will also lead to difficulties in attributing learning outcomes to use of the specific technology (Scanlon, Jones, Barnard, Thompson and Calder, 2000). Put forward as a negative feature of experiments, Gunn (1997) points out that the rigid nature of experimental design restricts the research. This limitation may however have its advantages when trying to interpret results. Despite differences between evaluations and experiments, similar measures may be used in both.

2.2.2 Process versus outcome

One approach to the classification of methods is to consider which aspect of the activity is evaluated. In relation to assessment, Heppell (2000) has argued for moving the focus from product to process. The way a student completes a task should be considered as important as the final product. This distinction is also made in studies that explore reading (Dillon, 1992; Schumacher and Waller, 1985). Process measures deal specifically with how readers use documents, and outcomes (or products) are reading rates and comprehension. Both process and outcome are appropriate to the evaluation of learning technologies and their use varies among studies.

2.2.3 Qualitative versus quantitative

Much is made of the 'paradigm debate' (Oliver, 2000), which concerns qualitative versus quantitative techniques. This debate will not be elaborated further as it has received sufficient attention by other authors. Fortunately not all authors of evaluation studies feel they need to take sides by adopting only one methodology (e.g. Woods and Keeler, 2001).

2.2.4 Subjective versus objective

A distinction in methods that is also relevant, but not given the same emphasis as the above debate, is the difference between subjective judgements and objective performance. Although the importance of measuring learners' perceptions of many aspects of VLEs should not be understated, such measurements cannot indicate, for example, ease of use nor ability to support learning. In an evaluation of VLEs and learners, Richardson (2001) explored whether individual differences of learners affect their perceptions of virtual learning environments. This is an extremely interesting research question. However, it would also be interesting to know whether individual differences affect learning performance.

In reflecting on the implementation and evaluation of two case studies on online interactivity, Boyle and Cook (2001) comment that student attitudes, obtained by questionnaires, do not indicate the quality of debate. However, marks from tutors for individual contributions (performance, albeit marked subjectively) and patterns of exchanges can provide useful information. As is often the case, employing different methods, hoping to converge on a single outcome, is a sensible policy. In exploring online teaching and learning materials in IT for art and design students, Brown, Hardaker and Higgett (2000) assessed their effects through questionnaires asking for student opinions and analysing their performance.

2.2.5 Expert versus user

When gathering subjective judgements, evaluations may adopt a technique from usability studies, heuristic evaluation, or ask for feedback from learners, as discussed above. In heuristic evaluations, a small number of 'usability experts' evaluate the interface against a set of heuristics. This method was used by interface design students to evaluate the usability of sites developed at another university using an online cooperative work environment (Collings and Pearce, 2002). Interestingly this study indicated that expertise

is required if using heuristics based on Nielsen (1994), which may be difficult for beginners in the field of HCI to understand. It is unlikely that this technique would be suitable for a summative evaluation of learning outcomes, although teachers are probably carrying out an informal version of this test when developing material for inclusion in a VLE.

2.3 Measures

A sample of measures are briefly described to illustrate different approaches. In general, what is measured determines the type of data that needs to be collected, the stage of activity to focus on, and who provides the data. The measures are chosen to answer the research question (in the case of an experiment) or provide the appropriate feedback in an evaluation. Issues of usability can be addressed by looking at responses to the system and eliciting perceptions. Learning is generally assessed through outcomes, but perceptions may again be informative. There may also be interactions between the usability of the system and the nature and extent of learning. Therefore comparing participation in discussions may contribute to assessing the role of the interface in the facilitation of learning.

2.3.1 Usability heuristics

This method is described in 2.2.5 and is distinguished from other measures by using an expert (or semi-expert) to conduct the evaluation. Although limited in many respects in comparison with other methods, this technique is efficient and may identify potential difficulties at an early stage without inconveniencing users. It may therefore be appropriate as an initial check before carrying out other sorts of evaluations.

2.3.2 Frequency of interactions

Jones et al. (2000) argue that interactions with the software are important to understanding the learning process. Logs of usage might include the use of resources and participation in discussion (Woods and Keeler, 2001; Holt et al., 2002).

2.3.3 Quality of interactions

Assessing frequency of contributions to discussions fails to differentiate between queries or comments, different topics (relevant or not), depth of debate, clarity of argument etc. If tools are employed and specific tasks carried out, it may be relevant to look at *how* these are used. Woods and Keeler (2001) report that dialogue accounted for 25% of the overall mark in the course they evaluated. This was graded on frequency, quality and timeliness. Judgements of quality are necessarily subjective, as are the majority of teachers' assessments (e.g. learning outcomes). Providing a set of criteria on which variables such as quality are judged can be helpful for future evaluations of this nature.

2.3.4 Learner perceptions

A range of variables can be measured by asking learners for their perceptions. Attitudes are sometimes separated out from perceptions (e.g. Jones et al., 2000), but essentially both are measured by asking for an opinion or judgement. It is the focus of the question that differs. This may be satisfaction, estimates of how much they have learned, usefulness of tools in the VLE, etc.

2.3.5 Learning outcomes

These are an essential measure of a VLE that supports learning, but there can be difficulties in interpreting the results. As mentioned in 2.2.1, it may not be possible to attribute changes in outcomes to specific elements of a learning technology. Nevertheless, studies may provide indicators of variables which may be important and these can provide the basis for future experiments.

The particular aspect of performance that is measured is determined by the objectives of the course, and is therefore likely to vary across studies. However, if measurement is limited to the defined objectives, the evaluation may fail to identify other incidental learning which may take place. Oliver (1997) introduces a dimension labelled 'domain independence' which relates to this distinction. He points out that learning outcomes can be related to the specific subject, or be more generic, e.g. organising discussion. There may also be subject-specific outcomes which are not specified or anticipated by the teacher, but would be worth identifying.

3. Summary of framework

The above discussion of the nature of evaluation is summarised in the following two tables. The framework is not intended to be exhaustive, but provides a method of positioning studies within the broad range of evaluations of VLEs that are conducted. Table 1 combines the purpose and methods of evaluation in the form of a matrix. Although the dimensions are broken down into distinct categories (i.e. evaluation or experiment,

process or outcome measures), studies may incorporate elements of each.

Table 1: Framework for distinguishing between evaluation studies based on their purpose and the methods that are used

			Purpose			
			Specificity/study design		*Discipline/focus*	
			Evaluation	Experiment	Usability	Learning
Methods	*Stages*	Process				
		Outcome				
	Type of data	Qualitative				
		Quantitative				
		Subjective				
		Objective				
	Participants	Expert				
		User				

The second table (Table 2) takes the sample of measures discussed in 2.3 and indicates which methods apply. The table can be read in this direction (i.e. down the columns) or across the rows to provide examples of measures which generate, for example qualitative, subjective data.

Table 2: Methods used in measuring specific aspects of VLEs

			Measures				
			Usability heuristics	Frequency of interactions	Quality of interactions	Learner perceptions	Learning outcomes
Methods	*Stages*	Process		✓	✓	✓	
		Outcome	✓			✓	✓
	Type of data	Qualitative	✓		✓	✓	
		Quantitative		✓		✓	✓
		Subjective	✓		✓	✓	✓
		Objective		✓			✓
	Participants	Expert	✓				
		User		✓	✓	✓	✓

4. Pilot study

4.1 Theoretical position

The pilot study is part of a research project based on a socio-cultural approach to cognition and consequently to learning. Learning is assumed to be socio-cultural in nature. Instead of electing an individual using a system as the focus of evaluation, the research considers that it is the social activity — inside which the system is used — that must be analysed.

Using the concept of *Legitimate Peripheral Participation* (Lave and Wënger, 1991), learning is defined as a consequence of members of a community engaging in a given activity. It is assumed that while engaged in the activity the group develops and incorporates knowledge. However, there must be a purpose or motive for such activity. Members take part in the activity because they have mutual objectives they believe will be achieved.

This description is extremely broad and applicable to many different types of social groups, even those not commonly related to any type of learning activity. Within the framework adopted learning occurs, by definition, in any place. Institutions such as schools and universities may be communities that are more specialized in teaching and learning, however learning is assumed to be part of any social practice.

A course may be construed as a social activity. Students and teachers are members of a group performing tasks to achieve their objectives. While engaged in the course members use artefacts such as books, journals, magazines, etc. to perform tasks. A VLE is one possible artefact that is available to the group. Attributes of the interface should be analysed to identify, for example, how efficient and satisfying the system is to use. However, it is insufficient to investigate how the VLE affects activity by only measuring its usability. Clearly, the evaluation activity incorporates various levels of analysis and variables such as ease of use should not be neglected. What

is required, in addition, is to incorporate variables that reflect the social-cultural component of the teaching/learning activity for which the VLE is intended.

4.2 Proposed variables

Since the learning activity is described in terms of Lave and Wënger's concept the research uses three of their variables. These variables are engagement, participation and achievement of goals. Table 3 defines the three variables in general terms.

Table 3: the three variables selected for use in the study

Engagement	The level of commitment a member has towards the activity, which is crucial for the establishment and achievement of objectives. It is measured by the number of utterances made during a chat session and the ratios number of utterances/number of participants and number of utterances/minute. It is also measured by the frequency with which a participant attends the chat sessions.
Participation	The range of different tasks performed during the activity. It is established by identifying different types of contributions made by the participants within the data.
Achievement of goals	The level of accomplishment of objectives and satisfaction with the activity. Questionnaires and course assignments are the source for the data.

The focus of the study is the social activity itself. This is defined as a group of individuals who join together to perform tasks to accomplish their objectives. By making the social activity the focus of the research, the evaluation of the VLE is placed within a context. The evaluation is no longer isolated from the situation in which it is used. It follows that making the system's use artificial by carrying out a controlled experiment is not desirable and the study should be as natural as possible. Obviously, any study is artificial at some level. Nevertheless, the pilot was conceived as an experiment using a real online course.

4.3 Method

The pilot uses two independent groups to compare two different VLEs with the same course content. The two systems are Blackboard and VirtusClass, developed at the Universidade Federal de Pernambuco (UFPE), Brazil.

A free online course (Basic Layout) was announced in e-lists of students and graduates from the graphic design programme of UFPE, Brazil. This generated more than 60 responses. After responding to enquiries and arranging the most convenient schedule, two groups were created with 13 students in each. Students used texts, PowerPoint presentations and links to sites. Seven chat sessions were held, during which the course material was discussed.

4.4 Outcomes

This pilot study tested the suitability of the variables. Both quantitative and qualitative data were collected. Unfortunately, a major fault during one VirtusClass chat session caused the loss of 4 out of 7 sessions. Nevertheless, the remaining data was used to inform the design of subsequent studies.

4.5 Engagement

There was a tendency for more engagement with BlackBoard. The *number of participants per session* and the ratio of *number of utterances/number of participants* and *number of utterances/minute* were consistently greater than in VirtusClass.

There may be two possible reasons for this. Firstly most participants were familiar with VirtusClass but not Blackboard and they may therefore have been more engaged by a novel environment. In particular, students found the whiteboard tool exciting. In the first meeting, one of the students asks about this tool[1]:

21	**SV >** I am curious. How will we use the above area?
28	**Lec >** This drawing tool is very cool, but I don't know yet if we will be able to use it.
29	**Lec >** the problem is always the same

[1] The number at the beginning of each line in the table represents the sequence of utterances within the chat session. In this extract the first mention of the whiteboard tool was made in the 21st utterance.

30	**Lec >** how to draw with a mouse
31	**LF >** that's terrific!!!!!
32	**SV >** it's easy, easy...

The second possible reason is the reliability of the BlackBoard system compared to VirtusClass. There were difficulties in connecting to VirtusClass which appeared to upset participants and they clearly got distressed. More than one participant may have dropped out of the course due to these problems.

4.6 Participation

Both environments have different sections where participants can perform specific actions. As a general rule, VirtusClass offers more options for student contributions while BlackBoard constrains them. For instance, in VirtusClass students are given the same status as instructors to post documents to everyone or suggest links to visit. Students can also initiate a discussion thread in the forum by themselves. It would therefore be expected that they would perform a greater number of roles in VirtusClass than BlackBoard. The only exception is the virtual classroom where BlackBoard's participants can use a set of tools — such as the whiteboard — not present in VirtusClass. The different types of participation students can perform in each section are listed in Table 4.

Table 4: Types of participation

Section	Type of participation student can perform	
	VirtusClass	*BlackBoard*
Link references	Access, Post, Search.	Access.
Document section	Access, Post.	Access.
Agenda	Access, Post.	Access.
Forum	Access, Post, Initiate thread.	Access, Post.
Virtual classroom	Chat.	Chat, Draw, Send particular message.

The virtual classroom and the forum seem to be special cases of participation. The dialogue inherent in these sections suggests participation should be analysed in qualitative terms. Stating viewpoints, supporting them, opposing other's perspectives are different types of participation in a dialogue that may be particularly relevant to the learning activity. For instance, the significance of using the whiteboard as a medium for student reasoning seems to be indicated by a passage from the Blackboard group.

After discussing the importance of grids for a layout, the lecturer tried to point out that they should not constrain the design. A grid, the lecturer sustained, can sometimes limit the layout possibilities. Although students kept saying that this was true, every time they rephrased their comments, they appeared not to get the point. This is that some ideas may not occur to designers if they are not conscious of how a grid may constrain the very conception of the layout. Then RS decided to draw the grid himself:

393	RS > well, I will draw then…
394	RS > it lacks the horizontal lines… etc… but on that grid one can make several different layouts

The lecturer reaffirms the advantage of grids and explains his point once more:

401	Lect > that's the great advantage of a grid, right?
402	Lect > it structures and permits variety
403	JB > and this is independent of its complexity...
404	Lect > what I was trying to highlight is that it may, only may, force constraints when not cleverly used.
	…
408	Lect > we must not allow the technique intended TO HELP
409	Lect > to become an OBSTACLE
410	RS > I see! Ok... it is great to keep these recommendations in mind to avoid stupid ideas when designing.

The lecturer then takes RS' drawing of a rectangular grid and draws a circle over it while RS repeats the lecturer's last sentence.

411	Lect > for instance…
412	RS > that's true. it must HELP, and never OBSTRUCT.
	...
415	Lect > all this square structure may very well lead us to avoid even thinking about a rounded picture
	…
419	Lect > notice that the circle may be seen as inside a square
420	RS > that's true.

421	Lect > but we tend not to think about a rounded picture while using a right-angle structure
	...
424	RS > I see... a matter of tendency
425	RS > like a newspaper ... we almost never see non-rectangle graphics.

RS finally seems to appreciate the point, and introduces a new idea about graphics in newspapers. This may be because RS and the lecturer changed their type of participation, which made RS think about the problem in different terms.

4.7 Achievement of goals

The lecturer did not require a formal task and the achievement of goals was restricted to students' impressions:

- "I found the document about calligraphy very interesting" (LL)
- "I really enjoyed the debate we had here" (EA)
- "I found the course incomplete… lacking a conclusion" (JB)

The pilot demonstrated the necessity of assessing the learning outcomes. It was decided to introduce a formal task — the design of a poster — into a second pilot. In addition, a questionnaire will be used to gather data such as personal satisfaction, particular difficulties and impressions about the learning achieved.

5. Conclusions

The pilot helped test the viability of the variables chosen for the study of VLE systems. It also produced valuable information for the design of the subsequent studies. The conclusions may be summarised as follows:

- The variables provide both qualitative and quantitative and objective and subjective data.
- Achievement of Goals needs to be assessed by practical tasks and a structured questionnaire.
- Technicalities such as connection reliability can severely hinder the experiment.

References

Boyle, T and Cook, J. (2001) "Online interactivity: best practice based on two case studies", *Association for Learning Technology Journal*, Vol 9, No 1, pp 94–102.

Brown, S, Hardaker, CHM and Higgett, NP. (2000) "Designs on the web: a case study of online learning for design students", *Association for Learning Technology Journal*, Vol 8, No 1, pp 30–40.

Chang, C-C. (2001) "A study on the evaluation and effectiveness analysis of web-based learning portfolio", *British Journal of Educational Technology*, Vol 32, No 4, pp 435–458.

Collings, P and Pearce, J. (2002) "Sharing designer and user perspectives of web site evaluation: a cross-campus collaborative learning experience", *British Journal of Educational Technology*, Vol 33, No 3, pp 267–278.

Dillon, A. (1992) "Reading from paper versus screens: a critical review of the empirical literature", *Ergonomics*, Vol 35, No 10, pp 1297–1326.

Draper, S,W, Brown, MI, Henderson, FP and McAteer, E. (1996) "Integrative evaluation: an emerging role for classroom studies of CAL", *Computers & Education*, Vol 26, No 1–3, pp 17–32.

Gunn, C. (1997) "CAL evaluation: future directions", *Association for Learning Technology Journal*, Vol 5, No 1, pp 40–47.

Heppell, S. (2000) "eLearning: how might eLearning really change educational policy and practice?" http://www.ultralab.ac.uk/papers/elearning/.

Holt, R, Oliver, M and McAvinia, C. (2002) "Using web-based support for campus-based open learning: lessons from a study in dental public health", *Association for Learning Technology Journal*, Vol 10, No 2, pp 51–62.

ISO 9241–11 (1998) "Ergonomic requirements for office work with visual display terminals (VDTs)–guidance on usability".

Jones, A, Barnard, J, Calder, J, Scanlon, E and Thompson, J. (2000) "Evaluating learning and teaching technologies in further education", *Association for Learning Technology Journal*, Vol 8, No 3, pp 56–66.

Lave, J and Wënger, E. (1991) *Situated Learning: legitimate peripheral participation,* Cambridge University Press, Cambridge.

Nielsen, J. (1993) *Usability Engineering*, Academic Press, San Diego, USA.

Nielsen, J. (1994) "Enhancing the explanatory power of usability heuristics" in *CHI 94 proceedings*, B Adelson, S Dumais and J Olson (Eds), ACM Press, Reading, MA, pp 152–158.

Oliver, M. (1997) *A framework for evaluating the use of educational technology*, University of North London, Learning and Teaching Innovation and Development (LaTID) London.

Oliver, M. (2000) "An introduction to the evaluation of learning technology", *Educational Technology & Society*, Vol 3, No 4, pp 20–30.

Preece, J, Rogers, Y, Sharp, H, Benyon, D, Holland, S and Carey, T. (1994) *Human-computer interaction*, Addison Wesley, Wokingham, England, p603.

Richardson, J. (2001) "An evaluation of virtual learning environments and their learners: do individual differences effect perception of virtual learning environments", *Interactive Educational Multimedia*, Vol 3, No, pp 38–52.

Scanlon, E, Jones, A, Barnard, J, Thompson, J and Calder, J. (2000) "Evaluating information and communication technologies for learning", *Educational Technology & Society*, Vol 3, No 4, pp 101–107.

Schumacher, GM and Waller, R. (1985) "Testing design alternatives: a comparison of procedures" in *Testing design alternatives: a comparison of procedures*, Thomas M Duffy and Robert Waller (Eds), Academic Press, Orlando, pp 377–403.

Woods, R and Keeler, J. (2001) "The effect of instructor's use of audio e-mail messages on student participation in and perceptions of online learning: a preliminary case study", *Open Learning*, Vol 16, No 3, pp 263–278.

