

Lam, P., & McNaught, C. (2004). Evaluating educational websites: A system for multiple websites at multiple universities. In L. Cantoni & C. McLoughlin (Eds). *ED-MEDIA 2004*, Proceedings of the 16th annual World Conference on Educational Multimedia, Hypermedia & Telecommunications, pp. 1066-1073. Lugano, Switzerland, 21-26 June. Norfolk VA: Association for the Advancement of Computers in Education.

Evaluating Educational Websites: A System for Multiple Websites at Multiple Universities

Paul Lam & Carmel McNaught
Centre for Learning Enhancement And Research
The Chinese University of Hong Kong
paul.lam@cuhk.edu.hk & carmel.mcnaught@cuhk.edu.hk

Abstract: The paper presents a strategy for systematic evaluation of 30 teacher-designed course websites in the year 2003, built under the project e3Learning which operates across three universities in Hong Kong. Because of the diversified nature of the websites, the evaluation system is designed to be able to facilitate tailor-made evaluation plans so that the evaluation data generated 1) allow in-depth understanding of the different types of course websites concerned, 2) provide the types of feedback the teachers request about their students' eLearning experience, and 3) enable the project team to better understand the nature of eLearning and thus provide better educational advice.

Background

University teachers are now more willing and able to use the Web to assist their teaching because the advance of technology has made the development of educational websites more convenient. The e3Learning (enrich, extend, evaluate learning; e3L) project has been designed to assist teachers to better exploit the possibilities of web-assisted teaching by offering a range of services from introducing teachers to practical ideas of using the Web in education, helping them better use the functions of teaching and learning platforms such as WebCT, putting up simple learning materials onto the Web, to developing complete course websites for the teachers. Full details of the design of this project are in James *et al.* (2003) and the project website <http://e3learning.edc.polyu.edu.hk/main.htm>. The e3L project operates across three universities, the Hong Kong Polytechnic University (PolyU), the City University of Hong Kong (CityU) and The Chinese University of Hong Kong (CUHK).

As more effort is spent on web-assisted teaching, the need to evaluate the various aspects of how the Web can help teaching and learning inevitably surfaces in order to allow us to better use our resources. Reeves & Hedberg (2003) suggests that the integration of evaluation into all technology-enhanced interactive learning systems is essential. This is the position taken by the e3L project. The project has an evaluation officer (first author) who actively assists teachers with evaluation of all the websites developed under the auspices of the e3L project.

Evaluating the web-assisted teaching of university teachers, though important, is difficult because of the highly diversified ways in which individual teachers are using the Web in their teaching. Effective and meaningful evaluation calls for a very flexible system in which different evaluation strategies are employed depending on the nature of the online project under investigation. The present paper highlights one such flexible system whereby 30 web-assisted teaching and learning cases were successfully evaluated in 2003. The overall design of the evaluation is a reflection-improvement model in which the findings of the evaluation contribute to further improvements in each of the web-assisted courses under investigation. There are evaluation resources already available, such as toolkits (e.g. Oliver *et al.*, 2002) or 'cookbooks' (e.g. Learning Technology Dissemination Initiative, 1998) but we chose to use a process mediated an evaluation officer in order to fairly rapidly build up a set of cases of good evaluation practice for Hong Kong university teachers to refer to. Our system (like all others) is not value-free and tends towards a naturalistic model (Guba & Lincoln, 1981; Alexander & Hedberg, 1994).

The process of evaluation consists of several stages:

- working out the **evaluation questions** with the teacher;
- choosing appropriate **strategies** to use in answering these questions;
- developing an **evaluation plan**;
- collecting **data** using the strategies selected, **analyzing** the data; and
- then making **decisions** about what to do next.

The findings lead to new decisions and actions on the part of both the teachers and the e3L project staff. Although each of the e3Learning websites go through the stages in this overall evaluation process, the exact evaluation design and schedule varies. Some of the factors that need to be considered are: 1) evaluation purposes, 2) the use of the Web, 3) evaluation questions, 4) evaluation data types, and 5) evaluation instruments.

Diversity in evaluation purposes

The exact evaluation strategies and schedule taken will be decided by the purposes of doing the evaluation. **Formative evaluation** and **effectiveness evaluation** are two main evaluation purposes. In formative evaluation, the purpose of the study is to elicit data concerning the particular use of the Web at an early stage of development in order to aid subsequent development. Effectiveness evaluation is done when the developed web elements are actually in use and looks at the effects the online experience have on the students and the teachers. We have decided to avoid the use of the term ‘summative evaluation’ as the reflection-improvement model employed in this study implies cycles of continuous improvement; there is no end to evaluation.

Formative evaluation usually calls for review-type evaluation strategies. The reviewers can be a selected group of student users, or a group of ‘content experts’ who are colleagues of the teachers; these reviewers then comment on the quality of the web content. Reviewers may be asked to fill in questionnaires, or may be invited to interviews or group meetings.

The schedule of effectiveness evaluation is very much constrained by the schedule of the course in which the web elements are used because the evaluation monitors the students’ engagement with the web-based elements. The variety of evaluation questions and strategies that are possible in effectiveness evaluation is explored below.

Diversity in the nature of web-assisted courses

The teaching and learning functions of the Web can be grouped into four main categories (McNaught, 2002). In reality, a course website may combine one or more of these functions. These different web functions, in turn, relate to different evaluation questions and different evaluation strategies.

Firstly, there are **content-rich** sites in which learning materials (plain text, graphics, pictures or multimedia) are put up to facilitate students’ learning. The evaluation questions to ask concerning this type of site are usually about the perceived usefulness of the learning materials and the learning improvement the materials are able to bring. In Hong Kong, this is the predominant type of educational website.

Secondly, there are **communication-rich** sites in which the teachers hold online discussion activities (through email, forums, chat-rooms, or video-conferences). The evaluation questions asked usually concern the quality of the communication, and the benefits perceived by students and teachers that communication brings to learning.

Thirdly, there are **assessment-rich** sites in which the teachers put up online quizzes and exercises for assessing students’ learning, or for drill and practice of basic concepts. The evaluation questions of this type of site mainly rest on the perceived usefulness of the quizzes, and the accuracy and fairness of the online quizzes if they are used for formal assessment purposes.

Lastly, there are sites that aims at providing **support functions** for teaching and learning such as providing course information to students, making online course-related announcements, and giving help on learning tips and learning

skills, etc. Evaluation questions concerning these functions usually relate to whether these support strategies are accessed and are perceived to be useful.

Diversity in evaluation questions

As mentioned in the previous section, the evaluation questions we want to ask concerning a particular web experience vary in accordance with the design purposes of the web elements. The evaluation questions also vary according to the focus the evaluation has on the various stages of the learning experience. Figure 1 illustrates various aspects of students' learning experience (Bain, 1999). Evaluation can focus on any or all of these aspects.

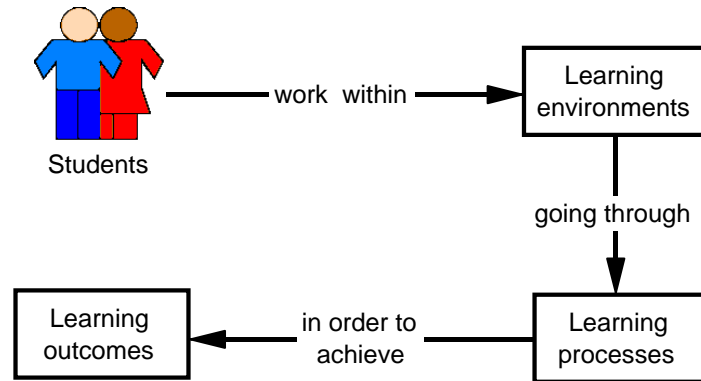


Figure 1: Stages of the learning experience

The evaluation can focus on one or more of these stages of the learning experience. For example, if we want to know whether students find a particular content-rich site valuable, there is a focus on the learning **environment** the students are experiencing. Evaluation questions like the following might be appropriate. Is the navigation clear? Do the resources download easily? Are these resources appropriate for the desired learning outcomes? Is the number of resources about right? Etc. If we ask how often and when the students complete the online quizzes and exercises in an assessment-rich site, we are considering the learning **processes**. Lastly, if we ask whether students are able to demonstrate critical-thinking skills in some online debate activities held in the forum of a communication-rich site, the focus is on the learning **outcomes** of the students' learning experience.

Diversity in evaluation data types

Different types of data are collected to answer the evaluation questions sorted out for any particular evaluation of the use of the Web. As Figure 3 below shows, there are '**feel**', '**know**' and '**do**' types of data. Data about what *students feel* are student perceptions (on the learning environments, on their learning processes, and even on how well they consider they have learned). A study by Williams (2002) indicates that the value that students place on online learning can be at variance with the expectations of their teachers, and so collection of this 'soft' data is very important. What *students do* when they study online provides data to help understand the learning processes. The data about what *students know* (e.g. through investigating the students' performance in assessments) may provide evidence to support that learning improvement has taken place through the particular eLearning experience under question and relates to the learning outcomes.

What the *teachers feel* is also important reflective data to help understand the observed learning environments and learning outcomes. Lastly, what the *experts feel* (and can more *objectively assess*) provides benchmarking data to help evaluate the quality of the learning environments of the site.

It is obvious that not all data types are needed in an evaluation. A flexible evaluation system enables the teacher, together with the evaluator, to design a coherent evaluation plan.

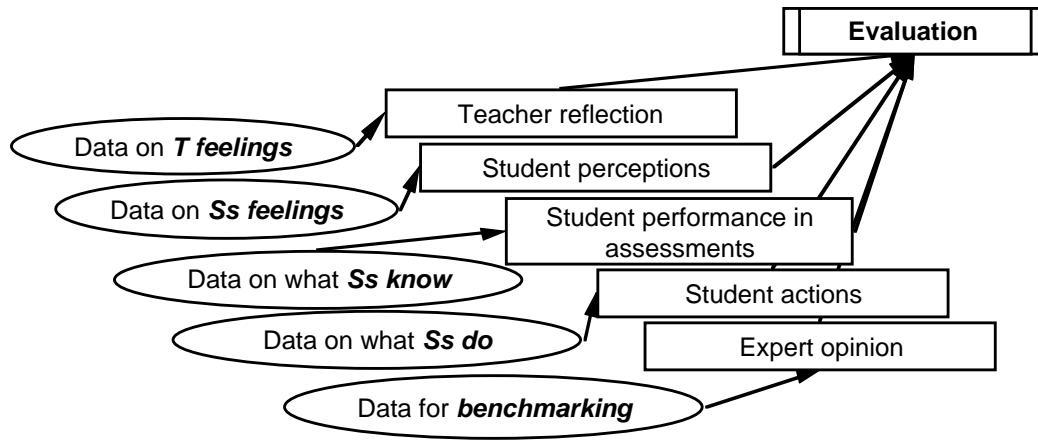


Figure 2: Evaluation data types

Diversity in evaluation instruments

Many evaluation instruments can be used to collect evaluation data of the above-mentioned types. Figure 3 shows that, for example, individual interviews, group interviews, satisfaction questionnaires, and Study Process Questionnaire (SPQ; measures students' approaches to learning; Biggs, Kember & Leung, 2001) are possible instruments to use to obtain data on what students feel (students' perception) about the site. To tap into what students do on the site, we may put up site access counters and even organize (video-taped) observation sessions in which we observe what students actually do when they work in the intended site. Students' examination and test performances, essays, assignments, forum postings and reports are good data to tap into what students' know. To understand what teachers' feel, we may ask teachers to write reflective journals and attend interviews with the evaluators to talk about their feelings towards using the site in teaching. Lastly, interviews, checklists, and reviewers' reports are common instruments to use when we elicit feelings from expert reviewers (who may be experienced teachers or eLearning specialists).

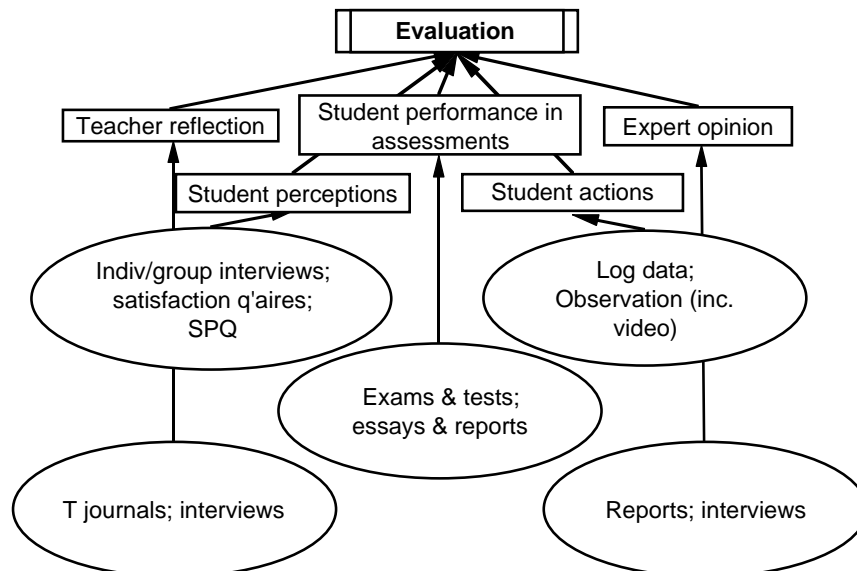


Figure 3: Evaluation instruments

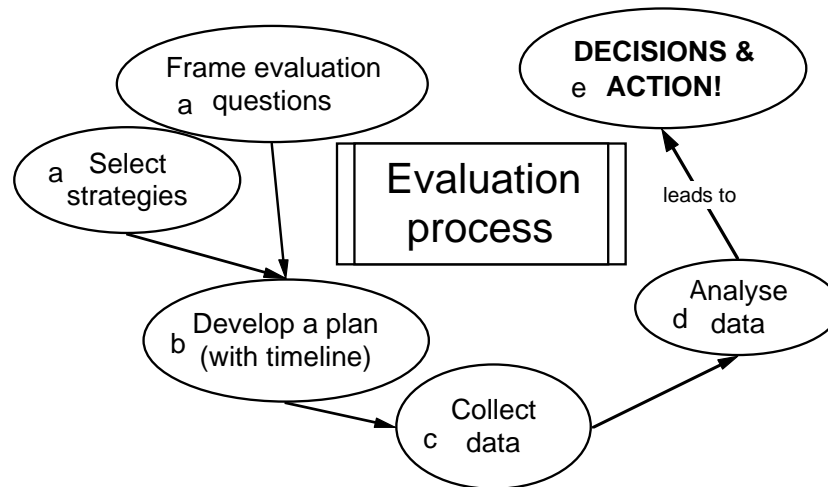
It is obvious also that we cannot and need not use all the instruments in any single evaluation (and there are other strategies; e.g. see <http://e3learning.edc.polyu.edu.hk/strategies.htm>). The strategy is to pick the best instruments

that can provide the maximum amount of information permitted by the constraints of time, resources and the availability of the students and teachers.

The mechanism

It becomes clear now that our evaluation system needs to be flexible so that it is able to provide tailor-made evaluation designs as a result of the diversity that exists in the domains of evaluation purposes, website nature, evaluation questions to be asked, data types that can be collected, and evaluation instruments that can be used.

In order to cope with this diversity, it is the main feature of our evaluation mechanism that we cooperate closely with the teachers from the beginning to the end of the whole evaluation process. Figure 4 serves to highlight the activities we have with the teachers and students throughout the evaluation.



<i>Evaluation activities</i>	
a	<ul style="list-style-type: none"> • Meeting with teachers. • Getting a clear picture of how the Web is used. • Deciding the evaluation questions. • Deciding the types of data to collect and the instruments to use.
b	<ul style="list-style-type: none"> • Putting the decisions into an evaluation plan.
c	<ul style="list-style-type: none"> • The evaluation officer designs the various evaluation instruments needed . • Teachers help to finalize the instruments. • Teachers help to administer them to the class. • The evaluation officer collects the feedback.
d	<ul style="list-style-type: none"> • The evaluation officer analyzes the feedback. • The evaluation officer writes reports.
e	<ul style="list-style-type: none"> • Reports are handed over to the teachers. • Meeting with teachers to look at the reports together and think about follow-up actions. • The evaluation officer and supervisor conduct meta-analyses of the data collected across several projects.

Figure 4: The evaluation activities in the different stages of the e3L evaluation process

As indicated in Figure 4, we meet with the teachers in the very beginning of the evaluation process. In this meeting, we understand the sort of assistance the teachers are going to introduce to the class. We suggest evaluation questions based on the nature of their website and also listen to the other evaluation needs of the teachers. Together, we decide the types of data to collect and the instruments to use taking into consideration limitations such as the availability of

the students and the teachers. We also set the time schedule for the carrying out of each of the selected instruments. After the meeting, we put the decisions concerning evaluation questions, data types, evaluation instruments, and the evaluation schedule arrived at into an evaluation plan. The plan serves to guide the evaluation work and procedure. (See http://e3learning.edc.polyu.edu.hk/evaluate_scenarios2_plan.htm for a sample evaluation plan.)

The evaluation officer designs the various evaluation instruments needed based on the evaluation questions on the evaluation plan. Teachers help to finalize the instruments and administer them to the class. The evaluation officer starts to collect the feedback according to the timing set down on the evaluation plan. Then, he analyzes the feedback (see http://e3learning.edc.polyu.edu.hk/evaluate_scenarios2_counter.htm for some samples of data) and writes reports. These reports are handed over to the teachers at the end of the evaluation process. (See http://e3learning.edc.polyu.edu.hk/evaluate_scenarios1_studpreport.htm for an abbreviated formative evaluation report.) Meetings are held to help teachers understand the reports and take action to improve the eLearning design based on the ideas in the reports. The e3L project also gets a deeper understanding of the issues in eLearning by conducting meta-analyses of the data collected across various websites.

Advantages of the system

Ash (2000) remarked that efficient and cost-effective evaluation has the following four characteristics:

- The evaluation must be undertaken with specific use and users in mind.
- Evaluation must be an activity that is integral to university operation.
- Evaluation must be cost-aware.
- Evaluation must be situation and context aware.

The present evaluation system can be considered cost-effective as it meets the above four criteria. First, there are two immediate and very clearly-defined stakeholders of the evaluation: the teacher who needs the evaluation in order to learn what worked and what didn't, and the e3Learning project team that needs the evaluation to learn what kinds of eLearning designs worked and what didn't work so that the project team can use the knowledge to build better eLearning materials for other teachers in the future.

Secondly, although the evaluation is not integral to our universities' operations, it is a required feature of the e3Learning project so that all teachers who ask for the project's help in developing web-based materials are involved with evaluation as well. We acknowledge that most teachers who come to the e3Learning project for help put developing the materials as their first priority, and do not find evaluation as immediately important. As Reeves & Hedberg (2003) suggested, "... evaluation is often overlooked or shortchanged in the haste to generate a product and deliver it on time" (p. xi). It is therefore a strength of the present evaluation system that the evaluation component is a required step of the whole e3Learning service. Otherwise, a very significant number of the teachers would not choose to, or have the time and experience to, carry out quality evaluation. However, in a quality-assurance-driven climate, teachers are using our evaluation reports in teaching portfolios, and we are hopeful that our evaluation strategies will have increasing impact in our universities.

Thirdly, our evaluation is cost-aware because of the fact that all the evaluation activities are done by staff who work full-time on website evaluation. This saves more time and hence more money than a decentralized system where the teachers are each responsible for their own sites. The evaluation services carried out for the 30 evaluation projects reported in this paper were effectively carried out by one postdoctoral evaluation officer (who also works on other projects), with the assistance of three students helpers, who each worked no more than three hours a week.

Fourthly, the evaluation conducted is extremely situation and context aware. As mentioned in the previous section, the evaluation team tailor-make each evaluation plan based on the specific nature and requirements of the website. We also cooperate closely with the teachers so that we know the details of the actual use of the Web if it differs from the originally planned website. All the information about the situation and context will eventually go into the evaluation report, making the evaluation highly context-sensitive. The richness of background information improves significantly the explanatory power of the evaluation data collected. It also makes meta-analysis possible and desirable as the project can contrast and compare various uses of the Web that are either similar to or different from each other in one or two potentially important ways.

As well as being cost-effective, the present evaluation also allows for in-depth and quality investigation and it is believed that this evaluation mechanism is highly sustainable.

In-depth and quality investigation

The tailor-made feature of the present evaluation mechanism has resulted in highly individualized evaluation plans, evaluation questions and instruments. Compared with generic evaluation methods where standardised evaluation instruments are used with all of the diversified sites, it is easy to see that the present individualized approach makes possible an in-depth investigation of the various web functions which may not be common across all sites.

Table 1 helps to illustrate this point of the individualized approach in our evaluation system. It can be seen on the table that we carried out three formative evaluations in 2003, 10 effectiveness evaluations on content-rich sites, 11 on communication-rich sites, four on assessment-rich sites, and two others on websites with a support focus. Of course, the evaluations plans in the same category also differed from one another in several ways.

	Formative	Content-rich	Communication-rich	Assessment-rich	Support focus
Feb 03 – June 03	1	3	3	1	0
July 03 – August 03	0	1	1	0	0
Sept 03 – Jan 04	2	6	7	3	2
Total	3	10	11	4	2

Table 1: Diversity of the evaluations

The present evaluation system also makes possible the employment of varied data sources by the use of a varied collection of evaluation instruments. As mentioned above, the individualized approach of the evaluation system allows the evaluator to discuss with the teachers the appropriate data sources to collect and the instruments, normally with the intention to collect the maximum amount of information allowed in the situation. As a result, several instruments are employed, and the data collected range from what students feel, know and do to what teachers and expert reviewers feel, thus providing a rich source of information essential to quality investigation. Table 2 shows the diversity of the instruments used for our 30 evaluation cases in 2003.

	Surveys (students)				Surveys (teachers)	Focus groups	Interviews	Web logs	Students' Production	Others
	Course -start	Task-end	Mid-term	Course -end						
Feb 03– June 03	1	6	0	6	6	2	1	0	2	0
July 03–Aug 03	0	0	0	2	2	2	2	2	2	0
Sept 03–Jan 04	2	0	1	28	20	5	3	15	6	2
Total	3	6	1	36	28	9	6	17	10	2

Table 2: Diversity of the instruments used

Sustainability

The cost-effective nature of the present evaluation system, together with the fact that it tends to result in in-depth and high quality investigations, makes the system highly sustainable. Its sustainability is also going to be improved by the project's effort in further streamlining the evaluation activities by making some evaluation instruments

reusable. One of the goals is to produce survey templates and checklists for various common and major uses of the Web to shorten the instrument development time for new evaluation projects of a similar kind.

The project has also started to compile an evaluation question pool (which currently has 290 items organized under 29 headings). This is a summation of all the questions asked in our various individualized surveys so far. It is thought that the pool will serve to be a good starting point for evaluators to use in compiling a survey form in areas where survey templates are not available.

Furthermore, the project has also developed a flexible counter system to keep track of the site logs of the websites it is monitoring. The counter system records the dates and times of site access and it provides a user-friendly platform for the evaluator to constantly monitor multiple pages of multiple websites, further streamlining the evaluation work of the project.

Conclusion

Considering the diversified nature of how the Web is employed by university teachers to assist teaching and learning, the present paper proposes and describes an evaluation system which has the flexibility to carry out individualized evaluations for each educational website. It is argued that this kind of individualized approach will result in in-depth and quality investigation, and that the system, with its cost-effectiveness and the potential re-usability of many of the refined evaluation instruments, is highly sustainable. The evaluation system had its trial run in 2003, studying 30 websites built by the e3Learning project. The magnitude and the richness of the data collected for these 30 evaluations have already proved that the evaluation system is an appropriate direction in which to go. The evaluation team see that the coming year will be used to extend the evaluation service to a further round of projects and to conduct meta-analysis on the collected data.

Acknowledgements

Funding support from the University Grants Committee in Hong Kong is gratefully acknowledged, as is the collaborative support of many colleagues in our three universities.

References

- Alexander, S., & Hedberg, J. (1994). Evaluating technology-based learning: Which model? In K. Beattie, C. McNaught & S. Wills (Eds). *Multimedia in higher education: Designing for change in teaching and learning*, pp. 233-244. Amsterdam: Elsevier.
- Ash, C. (2000). Towards a new cost-aware evaluation framework. *Educational Technology & Society* 3(4), 126-131.
- Bain, J. D. (1999). Introduction. Special issue: Learning-centred evaluation of innovation in higher education. *Higher Education Research & Development*, 18(2), 165-172.
- Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71, 133-149
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass Publishers.
- James, J., McNaught, C., Csete, J., Hodgson, P., & Vogel, D. (2003). From MegaWeb to e³Learning: A model of support for university academics to effectively use the Web for teaching and learning. In D. Lassner & C. McNaught (Eds). *ED-MEDIA 2003*, Proceedings of the 15th annual World Conference on Educational Multimedia, Hypermedia & Telecommunications, pp. 3303-3310, Honolulu, Hawaii, USA, 23-28 June. Norfolk VA: Association for the Advancement of Computers in Education.
- Learning Technology Dissemination Initiative. (1998). Evaluation cookbook. Retrieved on 30 April 2004 from <http://www.icbl.hw.ac.uk/ltidi/cookbook/contents.html>
- McNaught, C. (2002). Adopting technology should mean adapting it to meet learning needs. *On The Horizon*, 10(4), 14-18.
- Oliver, M., McBean, J., Conole, G. & Harvey, J., (2002). Using a toolkit to support the evaluation of learning. *Journal of Computer Assisted Learning*, 18, 199-208.
- Reeves, T. C.; & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Educational Technology Publications, Englewood Cliffs: New Jersey.
- Williams, P. (2002). The learning Web: The development, implementation and evaluation of internet-based undergraduate materials for the teaching of key skills. *Active Learning in Higher Education* 3(1), 40-53.