

Evaluating e-Learning Resources

Paul Riddy and Karen Fill

University of Southampton

p.j.riddy@soton.ac.uk

ABSTRACT

MECA-ODL, an EU funded project, developed an evaluation methodology and online tool for assessing eLearning materials, modules and courses. This paper describes the MECA-ODL outcomes and compares them with a number of other schemes. The MECA-ODL approach embraces the complete lifecycle of an eLearning offering, from conception through development to delivery and evaluation. Formal reviews and hands-on workshops produced positive feedback, plus some suggestions for improvement. Informed by these and the reviews of other schemes, the paper proposes enhancements to the evaluation methodology and tool, which have potentially wider significance.

Keywords

eLearning, quality, evaluation, tools

INTRODUCTION

The proliferation of eLearning resources, materials, modules, courses and programmes, poses a significant challenge to teachers and learners faced with deciding which might meet their needs and be of sufficiently high quality. A European team from universities in England, Spain and Germany, and training organisations in Italy and Greece, worked together for over two years to produce a methodology for the evaluation of online open and distance learning materials and Internet based programmes of study. The expertise of the multi-disciplinary team members spanned academic and commercial course design, development, delivery, management and evaluation. The aims of this EU funded project, Methodology for the Analysis of Quality of Open and Distance Learning delivered via the Internet, (MECA-ODL) were to:

- provide guidance for improving the overall quality of eLearning courseware
- facilitate buying decisions for tutors, learning portal managers and learners
- support eLearning developers and suppliers in the quest for quality.

The project developed through four main stages:

1. Compilation of a compendium and review of the scope of existing ODL resources.
2. Development of an evaluation methodology, based on the review results and existing schemes.
3. Development and testing of the on-line evaluation tool.
4. External evaluation of the on-line tool.

The major project outputs were a methodological guide for the analysis of quality in eLearning and a software tool that can be used online to evaluate materials, modules, courses and programmes.

This paper discusses the processes and outcomes of the project, with reference to other work, standards, guidelines and offerings in the field. Some of these were discovered during the project, others subsequently. All contribute to an evolving understanding of how best to evaluate eLearning resources.

THE MECA-ODL PROJECT

Building on previous work (NEPTUNO, 1999; Szamalk 1999) the MECA-ODL project partners compiled a compendium of ODL offerings in their own and other European countries. The scope and quality of these were reviewed and the team deliberated on the underlying design, development and delivery approaches. They agreed to work on a methodology encompassing seven phases and spanning commercial and academic requirements:

- | | |
|---------------|--|
| 1. Conception | - considers the readiness of the organisation to engage with eLearning |
| 2. Analysis | - establishes eLearning needs and methodologies |
| 3. Design | - pedagogical design and development process |
| 4. Content | - implementation of design including selection of technologies |
| 5. Production | - building the application, content, development team, sequencing |
| 6. Delivery | - marketing, course team, management, evaluation, certification |

7. Evaluation - course review, and feedback/development cycle.

Different team members worked on each phase initially, producing a list and descriptions of quality criteria. Their suggestions were then discussed by the whole team and refined until all partners agreed them. Three categories of potential users of the project outputs were also identified:

1. Developers of eLearning modules / courses
2. Users of eLearning modules / courses (i.e. learners)
3. Resellers of eLearning courses (i.e. trainers / tutors).

The project team considered every criterion with respect to its relevance in a phase and to each user type. Each criterion was also weighted, on a scale of 1 to 5, to denote its relative importance in an overall evaluation. The fully defined set of criteria, over all phases, numbers 143, with sub-sets used for the specific requirements of developers, resellers and users (learners). These are recorded in the guide (see Figure 1) and implemented in the online tool.

Phase: I. - Conception Criteria						
Code	D	U	R	Criteria	Weight	Comments
I.1 Building e-learning strategy.				To build an e-learning strategy means to take into consideration and change not only issues of technology but also issues of learning effectiveness.		
I.1.1	X			The learning culture in the organisation has been considered.	5	The impact of learning culture on effectiveness should be considered.
I.1.2	X			The learning background and interests of the organisations' staff have been considered.	4	The impact of staff factors should be considered.

Figure 1: Example from the printed guide of the Conception Phase

Note: the D/ U / R columns indicate the user category the criteria apply to: Developer, User or Reseller.

The MECA-ODL online tool

The online tool, implemented in English, German, Greek, Italian and Spanish, is currently available via the Internet at <http://wipaed-dbase.sowi.uni-bamberg.de/eLearning/WebObjects/mecaODLtool>. At registration a user selects the appropriate category (developer, reseller or user) and is subsequently presented with the relevant phases and criteria when evaluating an eLearning offering. The evaluation process involves awarding a 'score', again on a scale of 1 to 5, for the implementation of criteria (see Figure 2).

Figure 2: Data entry screen for application evaluation

The user may choose to ignore or score each criterion presented and, optionally, to change the default weighting that the project team decided was appropriate. Evaluations can be partially completed and returned to at a later date. The tool calculates an overall score for each phase, displaying this as a number. Scores for all the criteria are also shown as numbers and graphically (bar charts) and are automatically compared to the MECA-ODL 'ideal' that has been derived from the weightings agreed by the project team. Examples are shown in Figure 3.

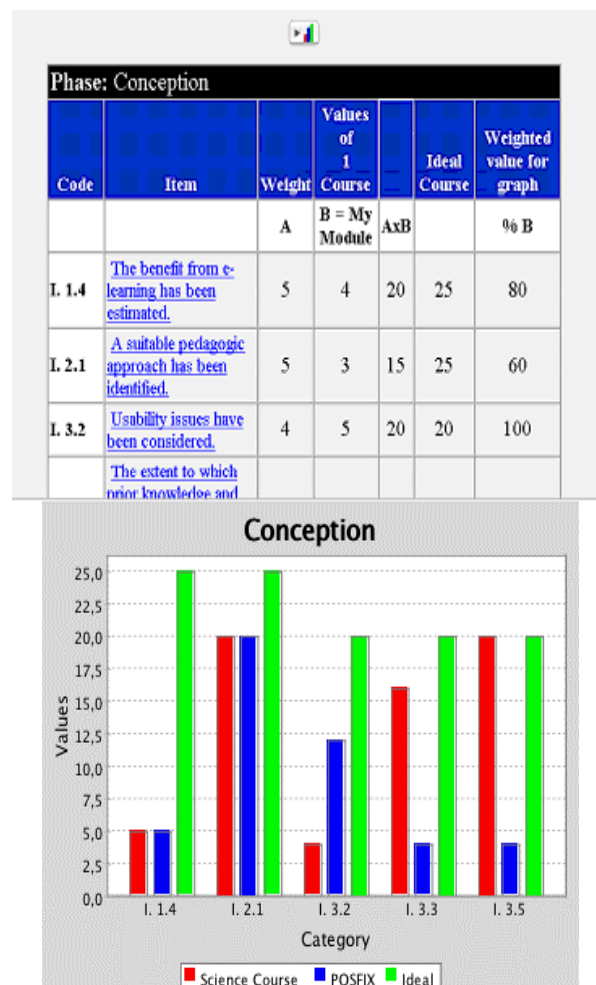


Figure 3: Numerical and graphical comparisons of quality

The tool was designed with flexibility for users to input their own set of weights for each of the criteria, but it is important that the same set is used if cross-comparisons between applications are required. At administrator level, the base set of criteria and reference weightings can be also be changed. The criteria, general descriptors, and operational instructions are stored in a database, making it straightforward to create a version in another language.

As mentioned above, the tool was designed to span the spectrum of eLearning evaluation including both commercial and Higher Education (HE). In reality, within HE the full range of criteria will be rarely used, as many of the criteria in the early phases are not considered, nor relevant, in the HE course selection process.

External assessment

The MECA-ODL guide and tool were assessed in late 2002 by eleven external evaluators, working in HE and commercial training, from all the countries involved in the project. Without exception they felt that an evaluation tool was needed. Common benefits of the MECA-ODL approach were deemed to be:

- the provision of a useful checklist for developing training/educational applications
- the graphical function for comparing different applications
- the concept of weighted criteria
- the flexibility to adapt for own needs.

Although some commented unfavourably on the large number of criteria, interested parties who tried out the tool at the first UK workshop suggested there are not too many if there is the flexibility to select appropriate sub-sets.

Observations and suggestions for further developments made by the external evaluators, plus participants in the two workshops the authors of this paper have run in the UK to date, include:

- Clearer language required
- Make it more user friendly
- Give more information on pedagogic models
- Glossary & FAQ's required Provide 'case study' examples of use
- Distinctive criteria sets for different purposes, e.g. commercial, academic
- On-line training course for evaluators
- Wizards – e.g. to set up an evaluation (criteria set)
- Repository of example on-line evaluations
- Link with learning objects metadata.

OTHER STANDARDS, GUIDELINES AND TOOLS

The main body involved with standards on a global scale is the International Organisation for Standardisation (ISO). This represents a hybrid of organisations from the public and private sector, with members from 147 countries, and its standards are truly international: "International Standards provide a reference framework, or a common technological language, between suppliers and their customers - which facilitates trade and the transfer of technology." (ISO, 2003). The ISO standards, which relate to this topic, are:

- **ISO 9000** - the ISO 9000 family of international quality management standards and guidelines has earned a global reputation as the basis for establishing quality management systems.
- **ISO 14915** - ISO 14915-1:2002 establishes design principles for multimedia user interfaces and provides a framework for handling the different considerations involved in their design. It addresses user interfaces for applications that incorporate, integrate and synchronize different media. This includes static media such as text, graphics or images, and dynamic media such as audio, animation, video or media related to other sensory modalities

Similar in concept, the Total Quality Management (TQM) model is widely used, but is not actually a standard. This represents a structured system for satisfying internal and external customers and suppliers by integrating the business environment, continuous improvement, and breakthroughs with development, improvement, and maintenance cycles while changing organizational culture. (IQD, 2001)

These, and similar schemes, are concerned with broad aspects of quality. They are complex, weighty and, for many educational developers, probably unusable. They enshrine notions of quality that start with development and continue beyond production, but are actually of little use for education.

Other bodies, working in the area of Learning Objects (LO's), have now produced tightly defined specifications and standards which consider the technical implementation, interaction (interoperability), and educational qualities of LO's. Two of the major international players here are the IMS Global Learning Consortium (www.imsglobal.org) and the Institute of Electrical and Electronics Engineers (www.ieee.org). Activities specific to the UK are coordinated by the Centre for Educational and Interoperability Standards (CETIS, 2003). These developments are already having an impact on the development of eLearning and future evaluation tools.

Evaluation of eLearning Resources

To date a number of evaluation schemes for the quality of eLearning resources have been developed. Unlike MECA-ODL, other examples discussed below do not cover all aspects of resource design and implementation.

E-Learning Courseware Certification (ECC)

Promoted by the American Society for Training & Development (ASTD), ECC deals with asynchronous Web-based and multimedia courses. ASTD supports its standards with examples, clarifications, definitions, scoring criteria, and other information. The ECC scheme evaluates compatibility, interface, production quality and instructional design of elearning courseware and now offers an on-line pre-evaluation tool:

which allows the end-user to pre-screen an asynchronous learning course against the 19 standards that are in place. The browser-based tool allows the end-user to open a course in one

window and the tool in another so that the course can be reviewed. The tool keeps score based on input from the end-user. (ASTD, 2002).

Final certification is done by external evaluators. ASTD is planning to offer an online database of certified courses.

Differences: The scope of the evaluation criteria does not engage with the readiness of an organisation to deliver eLearning, including the existing training culture. The criteria potentially offer a good assessment of other aspects of a course, but it is difficult to tell from the information available. ASTD offer a certification service for eLearning courseware, for which they charge a fee, and presumably have experts well versed in all aspects of course design and development to ensure comprehensive evaluation.

Quality on the Line (QoL)

The American Institute for Higher Education Policy reported on '24 benchmarks considered essential to ensuring excellence in Internet based distance learning' (IHEP, 2000). These were arrived at after a review of the literature and research with institutions delivering distance learning to determine which benchmarks were in use. The benchmarks are divided into seven categories of quality measures, namely:

- Institutional support
- Course development
- Teaching/learning process
- Course structure
- Student support
- Faculty support
- Evaluation and assessment.

Differences: This is a manageable set of criteria by which to evaluate the quality of a course but, in our view, does not provide a complete enough picture, especially in the areas of course design, content and production.

Consumer Based Quality Guidelines for Learning Technologies and Distance Education (CBQG)

This consumer guide is based on extensive research into the literature related to technology assisted (distance) learning. The quality of education and training is defined by what makes distance learning modules effective and efficient from a consumer's perspective. The criteria are presented as a series of 15 questions, the scope of which is illustrated below:

- Acquired content skills and knowledge should be relevant, transferable, specific ...
- Necessary learning skills are acquired for course completion, lifelong learning, self-directed learning
- Completion takes the form of credit or credentials that are recognised by professional accreditation bodies and other educational institutions
- Return on investment of the learner's time, finances and energy meets expectations for accessibility, objective benefits, effectiveness, efficiency, customer satisfaction.

Differences: As consumer guidelines these provide a good starting point for a potential learner to assess if a course is likely to provide what they are looking for, and indirectly suggest evaluation of the design, content and implementation. However, the consumer focus leads to the guidelines being output based, and they do not offer sufficient detail for a complete evaluation.

EPENDISI scheme

This evaluation scheme came out of the EPENDISI project run by the University of Macedonia, and is "concerned with both social and practical acceptability of hypermedia courseware. The authors (Elissavet & Economides, 2003) reviewed a number of approaches to evaluation in deriving their scheme, and their 124 basic criteria are grouped under the headings below:

- A) Evaluation of the content
- B) Organisation and Presentation of the content
 - 1) Pedagogical Parameters: Instructional theories - Curriculum, Structure, Learners Control, Adaptivity, Collaborative learning
 - 2) Design Factors: Interactivity, Navigation, Feedback, Screen Design
- C) Technical Support and update processes
- D) Evaluation of learning: Process of learning .

Each criteria is rated on a 5 part Likert scale, and the 124 criteria include 24 specific for web-based applications.

Differences: This extensive criteria set reviews the overall ability of the application to deliver good quality learning, but the scope of these does not engage with the readiness of an organisation to deliver eLearning. The criteria set include 34 focussing on practical implementation features such as screen design. Most of the criteria can be readily mapped into MECA-ODL phases but, with very different notation, interpretation is required. It is

not clear how the 'social acceptability' of a hypermedia application is represented, unless it is through 5 criteria associated with collaborative learning and related.

AT&T's Network Operations Education and Training (NOET) evaluation scheme

As reported by Pisik (1997), the NOET evaluation form is divided into five areas with a total of 68 statements. Each statement is scored on a 3-part scale, with Not Applicable as a fourth option. The five areas are:

- 1) Content and Instruction
- 2) Learners
- 3) Job Transfer
- 4) Design and Packaging
- 5) Operation.

Differences: With the largest number of criteria grouped into content, production and delivery the scheme has the potential to give a good indications of an applications ability to deliver learning within the scope of a topic. The scope of the criteria do not engage with the readiness of an organisation to deliver eLearning, including the existing training culture. The five criteria allocated to the job transfer area were difficult to map into the MECA scheme, with four being different to those in any other schemes.

Comparison of eLearning Evaluation schemes

A summary comparison of the above schemes to MECA-ODL is given in Table 1 below. Where appropriate, evaluation criteria are mapped into the MECA-ODL scheme. Some schemes use a standards or benchmarking approach, giving guidelines which evaluators use to produce a report, and in these cases we have indicated the areas covered by the guidelines in the table.

MECA-ODL Categories	MECA-ODL	ECC	QoL	CBQG	EPENDISI	NOET
Conception	26		√		0	1
Analysis	18				0	0
Design	10	√	√		24	7
Content	19		√	√	40	34
Production	23	√	√		30	14
Delivery	31		√	√	26	12
Evaluation	16	√	√	√	4	0
Total criteria	143	19	24	15	124	68

Table 1: Comparison of eLearning Evaluation schemes

Note: √ - indicates one or more benchmarks fall within a MECA-ODL category

The table indicates the comprehensive scope of the MECA-ODL criteria compared to the other evaluation schemes. The descriptions above also highlighted the differences in terminology and organisation of evaluation criteria into categories. For example, criteria in the QoL *course development* category fall into several MECA-ODL categories. The CBQG are all framed as evaluative questions, but are equivalent to some of the MECA-ODL criteria.

The standards and/or benchmark based schemes do not provide an indication of the importance of the criteria (weightings) and MECA-ODL is the only scheme with different sets of criteria for developers, users and re-sellers. These become particularly useful in the on-line implementation of the methodology, the evaluation tool.

Inevitably most of the schemes have a bias relating to their origins, and there are others which relate to specific topics or types of resources. For example, the MERLOT repository of Learning Materials provides a scheme for peer review. This contains guidelines for evaluation in the areas of Quality of Content, Effectiveness as a Teaching and Learning tool, and a check list for evaluating Ease of use. As with the other standards based schemes described above, the reviewer writes a report, rather than producing a print out of an evaluation table.

With the exception of CBQG, it would be difficult for those with little experience of on-line applications to complete an evaluation. This might seem an indictment of the schemes, but actually reflects the complexity of the task. Quoting from the MERLOT site for evaluating the "Potential Effectiveness as a Teaching-Learning Tool (MERLOT 2003)":

WARNING: This evaluation is the most difficult. Determining actual effectiveness requires actual use of the instructional software by real students and faculty. Evaluating POTENTIAL effectiveness is asking you to judge, based on your expertise as a teacher, if the instructional software is likely to improve teaching and learning given the ways the faculty and students could use the tool.

PROPOSED DEVELOPMENTS

Evaluation of eLearning resources requires a broad skill set. Users without any background in eLearning development and implementation would find it difficult to carry out an evaluation. Making a lot more information available via the user guide and on-line help system would go some way towards alleviating this problem. Ideally would-be-users would complete a module on eLearning evaluation to support use of the tool.

Proposed developments would remove some ambiguities in the criteria and produce revised sets of criteria for different user groups. For example, some of the other evaluation schemes are much more focused on the evaluation of *Content* and *Fitness for supporting Learning*. These are of major concern in Higher Education, and a set of criteria which supported these, plus relevant implementation and delivery issues would be widely used in HE. In conjunction with reduced criteria sets it would be helpful to have a simple way to set up the criteria and weightings from a particular user perspective. It is anticipated this could be done through developing 'wizards' which select the criteria set according to data input by the user. For example, one such wizard could provide a much-reduced criteria set for the novice eLearning evaluator.

Developments in the field of Learning Objects are creating a demand for more criteria evaluating the shape and interoperability of a LO, paralleling the language of Learning Design for developers, and simplifying it for users. None of the schemes reviewed explicitly consider such parameters and metadata. The MECA-ODL scheme contains a number of criteria concerned with LO's, which could be easily expanded and made more relevant. Any evaluation scheme or tool needs to be able to adapt to the rapid developments in the design and use of on-line learning resources, to be sustainable into the future.

ACKNOWLEDGMENTS

The MECA-ODL project was carried out with the support of the European Community in the framework of the SOCRATES programme. The content of this project does not necessarily reflect the position of the European Community, nor does it involve any responsibility on the part of the European Community.

REFERENCES

- ASTD(2002)Pre-Screening Tool Helps With E-Learning Courseware Certification.
http://www.astd.org/ecertification/self_assess.htm
- CETIS (2003) What is CETIS? <http://www.cetis.ac.uk/static/about.html>
- Elissavet G, & Economides A. A., (2003). An Evaluation Instrument for Hypermedia Courseware. *Educational Technology & Society*, 6(2), 31-44
- IHEP (2000) Quality On The Line: Benchmarks for Success in Internet-Based Distance Education.
<http://www.ihep.com/Pubs/PDF/Quality.pdf>
- IQD (2001) Definition of Total Quality Management. http://www.iqd.com/hoshin_def.htm
- ISO(2003)What'international standardization' means.
<http://www.iso.ch/iso/en/aboutiso/introduction/index.html#two>
- MERLOT(2003)Evaluation Criteria for MERLOT Learning Materials.
http://taste.merlot.org/projects/peer_review/criteria.php
- Neptuno (1999) Network of Entities promoting training in Universities with ODL. EU SOCRATES ref number: 25116-CP-1-96-1-ES-ODL <http://www.adeit.uv.es/neptuno/>
- Pisik, G. B, (1997). Is this course instructionally Sound? A guide to Evaluating Online Training Courses. *Educational Technology*, July-August 1997, 50-59.
- Szamalk(1999) Quality Assurance of Open and Distance Learning.
<http://www.szamalk.hu/okk/Projects/P010/Main.htm>