

Implementation of Quality Evaluation for Web-based Courses and Digital Learning Resources

Xia Teng^{1,2}, Brandon Muramatsu¹
Jian Wei Zhang³, Joseph G. Tront, Professor of Electrical Engineering⁴, Flora McMartin,
Director of Member Services and Evaluation⁵
Alice Agogino, Professor of Mechanical Engineering¹

¹ NEEDS, University of California, Berkeley, 3115 Etcheverry Hall, Berkeley, California,
94720-1750 USA

xia_teng@yahoo.com, mura@needs.org, agogino@needs.org

² Chinese E-Learning Technology Standards Committee, Beijing, China
zhangjw@tsinghua.edu.cn

³ Institute of Educational Technology, TsingHua University, Beijing, China

⁴ Virginia Polytechnic Institute and State University, 359 Durham Hall, Blacksburg, Vir-
ginia, 24061-0111 USA
jgtront@vt.edu

⁵ MERLOT, California State University, Long Beach, 1250 Bellflower Blvd., PSY-100
Long Beach, California, 90840-0901 USA
mcmartin@merlot.org

Abstract. The emergence of the Web as a delivery mechanism for education has led a number of organizations to develop and implement quality evaluation criteria for digital learning resources and Web-based courses. The Chinese E-Learning Technology Standardization Committee is developing a specification for evaluating Web-based courses. This paper provides a brief background on the standard, CELTS-22, and its guidelines for use. Further, this paper explores critical implementation issues through a case study of two similar evaluation criteria and systems that are used by NEEDS—A Digital Library for Engineering Education and MERLOT—the Multimedia Educational Resource for Learning and Online Teaching in the United States.

1 CELTS-22 Specification

The Chinese E-Learning Technology Standardization Committee (CELTSC) is chartered by the Chinese Information Technology Standards Committee and sponsored by the Chinese government (www.celtsc.edu.cn). Its goal is to develop standards for enabling interoperability and reusability of e-learning technologies, and for managing the quality of educational services. It is also responsible for developing the compliance test software for content delivery platforms and authoring tools.

The CELTS-22 specification describes a framework of quality characteristics of Web-based courses and defines corresponding evaluation criteria [1]. Following a standards development process, a consensus building phase was carried out using the Delphi Method, which is used to collect and distill knowledge from a group of experts in the design, development and management of instructional technologies [2]. The specification provides guiding principles for evaluating Web-based courses for various constituencies including:

- Providers: such as content, delivery platform and authoring tool vendors that design and develop Web-based courses.
- Consumers: such as organizations, institutions and individuals to select, acquire and use Web-based courses.
- Third party organization: such as organizations that evaluate, support or maintain Web-based courses.

The CELTS-22 specification is intended for Web-based asynchronous courses that can be taken independently. Typically these courses are a relatively complete learning experience that includes structured content with specified learning objectives, interactive learning activities and assessment of learning outcomes. The specification evaluates the potential of the contents and features of a Web-based course, not specific instances of the course. It is worth noting that elements of the CELTS-22 specification can be used to evaluate courseware or other digital learning resources that are used to enhance traditional or Web-based courses.

1.1 Quality Framework and Evaluation Criteria

The CELTS-22 framework defines quality characteristics for a Web-based course along four dimensions. Table 1 shows the dimensions (Content, Instructional, Interface and Technical Design) and whether the element is Mandatory (M) or Optional (O). In defining these dimensions and elements, CELTS-22 attempts to minimize overlap between each dimension.

Table 1. CELTS-22 Quality Characteristics

<p>Content Design</p> <p>1.1 Course Description (M) 1.2 Content-Objective Consistency (M) 1.3 Academic Quality (M) 1.4 Defining Content Object (M) 1.5 Content Sequencing (O) 1.6 Hyperlinks (O) 1.7 Resources Extension (O)</p>	<p>Instructional Design</p> <p>2.1 Learning Objectives (M) 2.2 Learner Control (M) 2.3 Learner-Content Interactivity (M) 2.4 Communication & Collaboration (O) 2.5 Motivation and Attention (O) 2.6 Presentation and Demonstration (M) 2.7 Media Use (M) 2.8 Learning Guidance (O) 2.9 Practice with Feedback (M) 2.10 Tracking (O) 2.11 Assessment (M)</p>
<p>Interface Design</p> <p>3.1 Style Consistency (O) 3.2 Layout (O) 3.3 Legibility (M) 3.4 Navigation and Orientation (M) 3.5 Links Labels (O) 3.6 Electronic Bookmarks (O) 3.7 Content Search (O) 3.8 Responsiveness (O) 3.9 Operational Help (M)</p>	<p>Technology Design</p> <p>4.1 System Requirements (M) 4.2 Installation and Uninstallation (M) 4.3 Reliability (M) 4.4 Multimedia Technology (M) 4.5 Compatibility (O)</p>

The framework is transformed into evaluation criteria by adding a description of the criteria, an expanded description and criteria indicators and a note with samples of the criteria. A descriptive sample of the evaluation criteria is provided in Table 2.

Table 2. Sample CELTS-22 Evaluation Criteria

Index	Name	Type	Description	Explanation	Notes
2.1	Learning Objectives	Mandatory	Clear and concrete learning objectives of each learning unit; higher level learning objectives of the main learning unit.	<p>Description: There should be a description of learning objectives for each learning unit.</p> <p>Clarity: The description should be concrete so as to gain certain knowledge, skill, or practical solution strategy, etc.</p> <p>Levels: Higher level learning objectives for the main learning units includes: analysis, summarization, evaluation or application of new knowledge for practical problem solving.</p>	A learning unit is a module in a course, such as a chapter, or a module defined in SCORM.

1.2 Use Guidelines

The CELTS-22 specification provides use guidelines for evaluating Web-based courses. Potential implementers of the specification should:

- Perform a requirement analysis to identify the needs and scope of the evaluation.
- Define a methodology for the evaluation to specify the appropriate evaluation criteria, data collection method and rating system.
- Conduct the evaluation to measure, analyze and present the results.

For Web-based courses to successfully complete an evaluation using the specification, they should:

1. Meet all the defined mandatory (M) criteria.
2. Achieve a minimum score for each dimension (where the actual score is dependent on the purpose of the evaluation).
3. Achieve a minimum overall score (where the actual score is dependent on the purpose of the evaluation).

2 Case Study: Implementing Evaluation Systems

This case study will examine the development and implementation of evaluation systems at the NEEDS—A Digital Library for Engineering Education ¹ (www.needs.org) and MERLOT—the Multimedia Educational Resource for Learning and Online Teaching² (www.merlot.org) digital libraries in the United States. Although the two examples in the case study focus on courseware and other digital learning resources, as opposed to the whole Web-based asynchronous courses of interest to CELTS-22, they are relevant examples because the criteria are similar (see Table 3) and they aim to accomplish similar goals.

Table 3. NEEDS *Premier Award* Evaluation Criteria³

Instructional Design	Software Design	Engineering Content
<ul style="list-style-type: none">• Learning Objectives• Interactivity• Cognition/conceptual change• Content• Multimedia Use• Instructional use/adaptability	<ul style="list-style-type: none">• Engagement• Learner interface and navigation• Technical reliability	<ul style="list-style-type: none">• Accuracy of content• Appropriateness of content

¹ NEEDS is an educational digital library where both educators and learners can search, access and download digital learning resources via the Web for engineering education [3].

² MERLOT is an international cooperative, based in the U.S., for high-quality online resources to improve learning and teaching within higher education.

³ Full details of the (2003) *Premier Award* evaluation criteria can be found at: www.needs.org/needs/public/premier/2003/2003-criteria-prelim.pdf.

2.1 Customize and Evolve the Evaluation Criteria

Both NEEDS and MERLOT have demonstrated the value in having discipline specific criteria that evolve over time. The evaluation criteria in use by NEEDS and MERLOT have been tested over seven years and five years, respectively. Both organizations recognized that to be successful, their evaluation criteria must be appropriate along a number of dimensions including the discipline and type of material being evaluated.

NEEDS, with the help of content experts, instructional designers and software designers, developed a set of high level categories (Instructional Design, Software Design and Content) and criteria that can be applied to courseware (or relatively large granularity learning objects) used by engineering educators [4-5]. Along with each criterion are statements (sub-components) to help users of the criteria determine how to evaluate the criterion (see Table 4). It is also worth pointing out that NEEDS reviews its evaluation criteria annually to determine how to maintain and improve the criteria's effectiveness and applicability for evaluating courseware. Major revisions were made after the criteria was first introduced in 1997 [6], and most recently in 2003 with a focus on Learning Objectives (see Sample Criteria in Table 4).

Table 4. Sample Criteria with Sub-Components

Category:	Instructional Design
Criteria:	1.1 Learning Objectives
Sub-Components:	<ul style="list-style-type: none"> • Learning objectives and goals are appropriate and clearly stated, in the software (preferred), in an instructor's guide or the submission packet. • The presentation and organization of content, as well as related activities, supports the learning objectives and goals. • Learners are aware of learning objectives as they are using the software and participating in the learning experience. • A clear method of measuring achievement of learning objectives and goals is provided within the software or by the learning experience. • Learning objectives and goals can be correlated to ABET⁴ accreditation criteria.

MERLOT followed a similar process of bringing together experts to draft a generic set of evaluation criteria⁵ applicable across the sciences and humanities in higher education in 1999. These criteria were then used as the basis for "custom" criteria used by each of MERLOT's discipline-based Editorial Boards to conduct peer review. Each discipline, therefore, has criteria customized to the nature of the particular discipline⁶.

⁴ U.S. Accreditation Board for Engineering and Technology (www.abet.org).

⁵ Full details of the generic MERLOT evaluation criteria can be found at: taste.merlot.org/projects/peer_review/criteria.php.

⁶ See the MERLOT-Physics criteria at taste.merlot.org/communities/physics/criteria.php and the MERLOT-History criteria at taste.merlot.org/communities/history/criteria.php for examples of "customized" criteria.

2.2 “Tiered” Evaluation

During the development of the evaluation criteria used by both NEEDS and MERLOT, it became evident that to be most effective the criteria would require a large investment of time and resources to do a thorough evaluation of digital learning resources. Both organizations balance the time and resource needs by implementing the “tiered” evaluation system described in Table 5. The “tiered” system of evaluation recognizes: (a) the need to develop processes that could scale with the ever increasing number of resources available on the Web and (b) the challenges and resources necessary to conduct detailed evaluations [4].

Table 5. Tiered Evaluation System used by NEEDS and MERLOT

Level	NEEDS	MERLOT
Base	NEEDS follows a collection development policy that encourages contributions of a wide range of materials [7].	MERLOT follows a collection development policy that encourages contributions of a wide range of materials. As resources are evaluated for whether or not they should be reviewed (“triage” process), some resources might be removed from the collection. In addition, in many disciplines, most of the resources in that disciplinary area have been reviewed to a lesser degree because they have passed that initial evaluation process.
Base+ Annotation	NEEDS encourages users to provide “Amazon.com”-style <i>User Comments</i> to further describe resources in its catalog.	MERLOT has a large number of both “Amazon.com”-style <i>Member Comments</i> and <i>Assignments</i> that provide context to the descriptive catalog records for each digital learning resource.

Level	NEEDS	MERLOT
Endorsed	NEEDS is collaborating with MERLOT-Engineering to develop, implement and conduct <i>Peer Reviews</i> of engineering resources.	MERLOT has focused on <i>Peer Review</i> using Editorial Boards with editors and reviewers that apply discipline-customized evaluation criteria. MERLOT is transitioning the <i>Peer Review</i> process from one that has been financially supported mainly by MERLOT partners, to that of a “professional” responsibility for educators in each disciplinary area that “volunteer” to serve as reviewers [8].
Premier ⁷	NEEDS developed the <i>Premier Award for Excellence in Engineering Education Courseware</i> to “recognize high-quality, non-commercial courseware designed to enhance engineering education.” The annual <i>Premier Award</i> is determined during a one and half day judging session by a panel consisting of content experts, instructional designers, students and publishers. The panel applies the evaluation criteria and examines the detailed documentation ⁸ to select one or more <i>Premier Courseware</i> of the year (and potentially one or more <i>Finalist Candidates</i>).	MERLOT has instituted the <i>MERLOT Awards Program for Exemplary Online Learning Resources</i> (taste.merlot.org/awards/) to recognize the “best” digital learning resources annually. MERLOT Editors select the “top” resource in each year and it is named the <i>MERLOT Classic</i> in that discipline. MERLOT editors then evaluate all of the <i>MERLOT Classics</i> to select one or more <i>MERLOT Editors’ Choice(s)</i> for the year.

2.3 Training of Reviewers

A critical element to the consistent application of evaluation criteria has been a structured training process for reviewers applying the criteria. The endorsed and premier levels of the “tiered” evaluation system require teams to evaluate the digital learning resources. In the case of MERLOT there are multiple teams of reviewers both within a given discipline and across disciplinary communities; in the case of NEEDS there is a single group of up to 10 reviewers for each annual judging panel. Goals in training reviewers include: (a) providing a shared understanding among the group of reviewers of the details and methods to apply the evaluation criteria and (b) “standardizing” the process to ensure inter-rater reliability (between individual reviewers and across multiple teams/panels).

⁷ NEEDS and MERLOT have agreed that NEEDS *Premier Courseware* are equivalent the *MERLOT Editors’ Choice* selections. In addition NEEDS *Premier Award Finalist Candidates* can be considered equivalent to *MERLOT Classics*.

⁸ 2004 *Premier Award* submission guidelines can be found at: www.needs.org/needs/public/premier/2004/submission/.

3 Summary and Conclusions

The CELTS-22 specification describes the general quality characteristics of Web-based courses and corresponding evaluation criteria for the Chinese educational community. The specification defines criteria and processes that can be used as a general guideline. A number of important issues can be identified from the brief case study of NEEDS and MERLOT, the importance of these issues should be recognized as the CELTSC works with organizations to implement the CELTS-22 specification. The three issues to consider when evaluating potential use of the CELTS-22 specification can be summarized as:

- Understanding that the specification will be customized and localized by implementers to meet their needs.
- Defining appropriate levels of evaluation (or levels of implementation) that balance resource and social constraints.
- Understanding that training is necessary to enable participants to evaluate Web-based courses using the specification.

References

1. China E-Learning Technology Standards Committee. Specification for Evaluating Web-Based Courses, CELTS-22.1 CD1.0. 2004. URL: www.celtsc.edu.cn
2. Teng, X., Duan, C., Zhang, J., and Wang, X. The Development of CELTS-22: the Standard for Evaluating Web-Based Courses. *Modern Educational Technology*, 11(1), pp .5-12, 2003
3. Muramatsu, B. and A. M. Agogino, NEEDS—The National Engineering Education Delivery System: A Digital Library for Engineering Education. *D-Lib Magazine*, 5(4), 1999. URL: www.dlib.org/dlib/april99/muramatsu/04muramatsu.html.
4. Eibeck, P. "Criteria for Peer-Review of Engineering Courseware on the NEEDS Database," *IEEE Transactions on Education*, Special Issue on the Special Issue on the Application of Information Technologies to Engineering and Science Education, 39(3), pp. 381-387, 1996.
5. Synthesis Coalition, Quality Workshop. Synthesis Coalition: San Luis Obispo, CA, 1995.
6. Muramatsu, B., P. A. Eibeck, J. L. Stern, and A. M. Agogino, "Effective Processes to Give Engineering Educators Easy Access to Quality-Reviewed Electronic Courseware," Invited Presentation, NSF Engineering Education Innovators' Conference, Washington, D.C., April 8, 1997.
7. Muramatsu, B. NEEDS Collection Development Guidelines. 2003.
8. Hanley, G. and F. McMartin. Scaling the Peer Review Process. NSDL All Projects Meeting, Washington, D.C., October 2003.