LEARNING OBJECT EVALUATION: COMPUTER-MEDIATED COLLABORATION AND INTER-RATER RELIABILITY

J. Vargo,* J.C. Nesbit,** K. Belfer,** and A. Archambault***

Abstract

Learning objects offer increased ability to share learning resources so that system-wide production costs can be reduced. But how can users select from a set of similar learning objects in a repository and be assured of quality? This article reviews recent developments in the establishment of learning object repositories and metadata standards, and presents a formative reliability analysis of an online, collaborative method for evaluating quality of learning objects. The method uses a 10-item Learning Object Review Instrument (LORI) within a Convergent Participation evaluation model that brings together instructional designers, media developers, and instructors. The inter-rater reliability analysis of 12 raters evaluating eight learning objects identified specific items in LORI that require further development. Overall, the collaborative process substantially increased the reliability and validity of aggregate learning object ratings. The study concludes with specific recommendations including changes to LORI items, a rater training process, and requirements for selecting an evaluation team.

Key Words

Learning objects, eLearning, collaborative, design, reliability, evaluation, Web-based education

1. Introduction

Developments in Web-based education and technology-mediated learning environments present educators with an important opportunity to increase students' access to secondary and tertiary education. At the same time, these developments offer potential for improving student learning through more widespread use of active learning strategies. While stimulating excitement among educational technology professionals, such developments stir fears of yet another resource-hungry enterprise, draining finance from already under-funded academic institutions. The nascent development of learning object standards and repositories offers a productive response to these fears. Sharing high-quality learning objects across the internet, developed by a few but used by many, enables cost-effective development and deployment of these expensive resources [1]. But how can educators be assured that the learning objects they find in online repositories are of high quality and can fulfil their objectives?

Systematic evaluation of learning objects must become a valued practice if the promise of ubiquitous, high quality Web-based education is to become a reality. The processes and tools adopted for learning object evaluation will need to efficiently balance the requirements for reliability and validity against time and cost. This article describes an evaluation process developed by the authors and provides evidence of its reliability. The context for this evaluation process is established by reviewing the characteristics of learning objects and repositories, and their role in Web-based education.

2. The Emergence of Learning Objects

With the explosive growth of the internet and the consequent increase in global connectedness, a new level of resource sharing has become possible. The online information revolution has spawned the learning object, the cyber equivalent of earlier shareable resources for education and training. Lecture handouts, textbooks, test questions, and presentation slides can all be considered learning objects. The online versions of these, together with interactive assignments, cases, models, virtual laboratory experiments, simulations, and many other electronic resources for education and training further add to the pool of learning object types. Many thousands of learning objects are now freely available through online repositories that can be searched using metadata that is being standardized by international and national organizations.

2.1 What is a Learning Object?

Knowledge element, learning resource, online material, and instructional component are all terms that have been used

* Department of Accountancy, Finance and Information Systems, University of Canterbury, Private Bag 4800, Christchurch, New Zealand; e-mail: john.vargo@canterbury.ac.nz
** Information Technology and Interactive Arts Program, Simon Fraser University, 2400 Central City, 10153 King George Hwy, Surrey, BC, V3T 2W1, Canada; e-mail: {nesbit, kbelfer}@sfu.ca
*** Microsoft Corporation, 1 Microsoft Way, Redmond, WA 98052, USA; e-mail: annea@microsoft.com

to mean much the same as "learning object". NETg, a major eLearning provider, defines a learning object as a resource with three parts:

(1) a learning objective,
(2) a learning activity, and
(3) a learning assessment [2].

However, taking a broader perspective on learning objects is the Learning Technology Standards Committee (LTSC) [3] of the Institute of Electrical and Electronic Engineers (IEEE). The LTSC is a standard-setting body with representatives from a wide range of organizations that develop products and specifications for eLearning. It defines a learning object as "any entity, digital, or non-digital, which can be used, reused, or referenced during technology supported learning" [4].

The diversity in types of learning objects is especially indicated by the three properties shown in Table 1: aggregation level, interactive type, and resource type. These properties are elements from the IEEE Learning Object Metadata (LOM) standard [5] that was approved in 2002. Referring to these properties, we can speak, for example, of a "level 1 expositive graph" or a "level 2 active simulation," but probably not a "level 4 expositive diagram."

Table 1
Types of Learning Objects Based on Elements
from the IEEE LOM Standard

Aggregation Level
Level 1 refers to the most granular or atomic level of aggregation, e.g. single images, segments of text, or video clips
Level 2 refers to a collection of atoms, e.g. an HTML document with some embedded images, or a lesson
Level 3 refers to a collection of level 2 objects, e.g. a set of HTML pages linked together by an index page, or a course
Level 4 refers to the largest level of granularity, e.g. a set of courses that lead to a certificate
Interactive Type
Expositive: information flows primarily from the object to the learner for and includes text, video and audio clips, graphics, and hypertext linked documents
Active: information flows from the object to the learner and from the learner to the object for learning-by-doing including, simulations and exercises of all sorts
Mixed: a combination of expositive and active
Resource Type
Resource types could include: exercise, simulation, questionnaire, diagram, figure, graph, index, slide, table, narrative, text, exam, experiment, problem, and self-assessment

But how can an educator, course designer or student find the appropriate learning object for their particular use among the profusion of online materials? Perhaps, the critical defining characteristic of the learning object concept is the ongoing development of a set of related metadata standards and specifications that permit learning resources to be searched for and retrieved in convenient and effective ways.

2.2 Metadata

Metadata can help users locate, license, develop, combine, install and maintain learning objects for online courses or programs. The IEEE LOM standard [5], drawing from earlier work by the IMS [6], ARIADNE [7] and Dublin Core [8] groups, specifies 59 metadata elements grouped into nine categories:

1. General
2. Lifecycle
3. Meta-metadata
4. Technical
5. Educational
6. Rights
7. Relation
8. Annotation
9. Classification

To meet the eLearning needs of the US Department of Defense, the Advanced Distributed Learning (ADL) Initiative [9] developed the Shareable Content Reference Model (SCORM) combining a range of technical specifications and standards including the IEEE LOM Standard. ADL has also developed conformance tests to verify that a learning object complies with specific aspects of SCORM, such as the metadata requirements. One can anticipate that SCORM will drive a substantial portion of eLearning providers to pack their current and future products with standard metadata.

Notably absent from the IEEE LOM standard are metadata on the quality of learning objects as judged by users or third-party assessors—the type of metadata with which this article is primarily concerned. We believe that the development of tools and formats for quality evaluation of learning objects will be the next major advance in LOM, and one that will have a powerful impact on the design of interactive media for Web-based education and training.

2.3 Growth in Size and Number of Repositories

Access to learning objects is typically gained via a repository that either maintains objects and metadata on a centralized server, or maintains metadata only and provides links to objects distributed throughout the internet. These repositories have grown in number, size, and sophistication since their inception in the mid-1990s. There appear to be four functional categories:

• Commercial repositories that offer access as a customer service to instructors and course developers. These include publishers' websites that provide instructors

Table 2
Examples of Open-Access Repositories

Repository	Year Founded	Items Indexed*	Description
Telecampus** http://telecampus.edu	1997	66,000	Online university courses and programs; mainly aggregation levels 3 and 4. No support for quality evaluation
Apple Learning Interchange http://ali.apple.com	1998	21,000	Online resources for K-12 education; aggregation levels 1 and 2. No support for quality evaluation
MathForum http://mathforum.org	1996	8,800	Online mathematics resources for K-12 and post-secondary education; mainly aggregation levels 1 and 2. No support for quality evaluation
Merlot http://merlot.org	1997	7,000	Online materials for post-secondary education (with some K-12); aggregation levels 1 and 2. Support for user comments and peer reviews
Alexandria/Careo http://belle.netera.ca	2001	2,500	Online materials for post-secondary education, mainly at aggregation levels 1 and 2. No support for quality evaluation
Harvey Project http://harveyproject.org	1999	600	Online materials and courses on human physiology; mainly university level; aggregation levels 1–3. Support for user comments and peer reviews
Wisconsin Online Resource Center http://wisc-online.com	1999	600	Centralized storage of online resources supporting Wisconsin's technical colleges; aggregation level 2. Support for user comments

* Shows the approximate number of items indexed in August 2002.
** Because the courses and programs it indexes usually require tuition payment and registration, Telecampus may not be regarded as an open-access repository.

who have ordered a textbook with related teaching resources such as slide presentations, cases, simulations, test item banks, and course content formatted for use in online course management systems.

- Corporate repositories maintained by commercial eLearning providers to support their own course development and delivery activities.
- Corporate repositories used by large companies and military organizations to train and develop internal personnel. Examples include Cisco, Honeywell, and American Express [10].
- Open-access repositories usually established by consortia of educational organizations. The central infrastructures for these are often funded with research or development grants, with the learning objects contributed by individual educators or participating institutions on distributed servers. Examples of this type of repository, mostly metadata repositories rather than centralized, object repositories, are provided in Table 2.

At the time of this writing, Merlot is the only repository of those listed in Table 2 that provides a quality-based sort of search results. Merlot returns object descriptions in descending order of quality rating, with non-evaluated objects listed last. The evident convenience of this feature is such that we expect quality-based sorting to become a common feature in future repositories.

2.4 Future of Learning Objects and Repositories

One can reasonably ask why open-access repositories are necessary at all, given the availability of easy-to-use and highly effective full-text web search engines. One answer is that potential users require metadata to retrieve non-text objects such as images or video, and they require standard metadata to identify an object as designed for learning and to efficiently select the best object to meet their pedagogical need. Repositories satisfy these requirements by providing tools for entering, storing, and retrieving object metadata. But to what extent do repositories need to be built on central databases? Recognizing that individuals and organizations vary in how they produce and use metadata, Hatala and Richards [11] have developed a peer-to-peer architecture and prototype in which standard LOM and quality reviews are globally distributed over a network of individual workstations and community or corporate servers.

A question raised by Wiley [12] is who or what should assemble learning objects into units and courses. Although one can imagine automated systems using repositories of level 2 objects to assemble level 3 objects, we are more likely to see widespread use of authoring systems that autonomously search repositories to recommend objects matching objectives and formats specified by a human author.

Taking an alternative perspective, many educators believe that students learn more effectively when, in

response to realistic problems, they construct their own knowledge by selecting resources and re-working them into different forms that they share with peers. Where this view prevails, we may see student teams searching repositories to assemble objects from which they can learn to solve complex problems. Objects developed by one team may be entered in the repository for use by other teams.

For the following reasons, these potential innovations are critically dependent on the availability of quality reviews, and on the people, processes, and tools that produce those reviews:

- Searching through peer-to-peer networks that link together multiple repositories and thousands of personal computers will yield many more hits than is presently obtained from any single repository, thus intensifying the need for quality-based sorting of search results.
- The prevalence of low-quality materials ensures that authoring tools automatically recommending resources will fail spectacularly unless their decision procedures include quality metrics.
- Hill and Hannafin [13] observed that often "students lack sufficient metacognitive awareness and comprehension monitoring skill to make effective [resource] choices." Because of the risk of students being misinformed by inaccurate content, or of wasting time with poor instructional designs, quality reviews become even more important in self-directed or learner-centred educational settings where students are expected to select their own learning resources. We note that evaluating resources is often regarded as an effective learning technique in itself; and we anticipate that students carrying out team-based development of learning objects would also benefit from collaboratively evaluating them.

3. Current Approaches to Learning Object Evaluation

Learning object evaluation is a relatively new problem with roots in, and overlap with, a substantial body of prior work on the evaluation of learning materials and courseware [14]. Primarily, it is the goals of sharing and reuse that determine how learning object evaluation differs from other evaluation approaches. Wiley [15] has pointed out that the reusability of a learning object is inversely related to its aggregation level. Thus, it is particularly the more reusable level 1, level 2, and to a lesser extent level 3 learning objects that require new evaluation approaches.

The Southern Region Educational Board maintains the EvaluTech repository [16] containing over 7,000 reviews of K-12 learning resources including books, CD-ROMs, videodisks, courseware, online courses, websites, and other level 2 and 3 materials. Separate evaluative criteria, with significant emphasis on content and often on technical features, are provided for these different media. The reviews provide no numerical rating that would allow quick comparison among resources or quality-based sorting of search results.

The American Society for Training and Development (ASTD) has developed a set of standards for certifying Web-based education courseware [17]. These include interface standards, compatibility standards, production quality standards, and instructional design standards. The instrument used in the present study substantially intersects with the ASTD standards, and in addition deals with qualities such as reusability, accessibility, and compliance with LOM standards.

The MERLOT site listed in Table 2 offers the best current example of mass application of learning object evaluation in Web-based education because it better supports evaluation of level 1 and level 2 objects. With comments and ratings on a five-point scale, MERLOT users and appointed peer reviewers evaluate three general properties: quality of content, potential effectiveness as a teaching–learning tool, and ease of use. The quality-based sorting of search results uses an equally weighted average of these three ratings. MERLOT's peer evaluation process is carried out by two subject-matter experts working asynchronously.

None of the learning object evaluation methods currently in use specify roles for a small team of reviewers with complementary knowledge in subject matter, instructional design, and media development. And, as of this writing, we have been unable to find previous reliability analyses of existing methods.

4. Collaborative Evaluation with LORI

The research reported in this article tested a process designed by the authors for the evaluation of learning objects. The process consists of two key components: the Learning Object Review Instrument (LORI) [18] that an individual evaluator can use to rate and comment on the quality of a learning object, and the Convergent Participation Model [19] that brings together a team of evaluators and their individual reviews to create and publish a collaborative LORI review. While LORI applies specifically to the evaluation of learning objects, Convergent Participation is a general evaluation model that could apply to a different domain when combined with an appropriate, domain-specific instrument.

4.1 The Structure of LORI

LORI (version 1.3) measures 10 separate qualities of learning objects:

1. Presentation: aesthetics
2. Presentation: design for learning
3. Accuracy of content
4. Support for learning goals
5. Motivation
6. Interaction: usability
7. Interaction: feedback and adaptation
8. Reusability
9. Metadata and interoperability compliance
10. Accessibility

The presence of each quality is measured with a single item using a rating scale consisting of five levels: absent (0), weak (1), moderate (2), strong (3), and perfect (4). A descriptive rubric is provided for each level. Efforts were made to design the scales to represent continuous dimensions with a similar distance between the levels. During this study and in earlier exploratory studies, users occasionally entered intermediate values between two adjacent levels suggesting that they perceived a continuous underlying construct.

4.2 The Convergent Participation Model

Convergent Participation is a two-cycle model designed to boost the efficiency and effectiveness of collaborative evaluation. In the first cycle, participants with diverse and complementary areas of expertise individually review a set of learning objects using LORI. The first cycle is completed asynchronously within a period of few days. In the second cycle, the participants come together in a moderated discussion using a synchronous conferencing system. During the discussion, participants adjust their individual evaluation in response to the arguments presented by others. At the end of the meeting, the moderator seeks consent of the participants to publish a team review synthesized from the mean ratings and aggregated comments.

When we began using LORI to evaluate objects within synchronous sessions, it became apparent that a large proportion of communicative acts were dedicated to exchanging ratings on individual items. There was often close agreement on some of the items, substantial disagreement on other items, and insufficient time to deal with all. A procedure was introduced whereby items are sequenced for discussion based on the level of inter-rater variation obtained from the first cycle: The items are ordered from lowest to highest variation and the moderator attempts to pace the session to cover all items on which there is substantive disagreement.

5. Research Goals of This Study

There were three research goals of the current study:

- To evaluate the inter-rater reliability of LORI when used as an assessment tool in a non-collaborative assessment setting.
- To investigate the use of LORI within the Convergent Participation model: How does collaborative assessment affect inter-rater reliability?
- To indicate needed improvements in LORI and the Convergent Participation model.

6. Participants

The participants were 12 adults with experience in the fields of educational technology, corporate training, or higher education; in all, eight educational technology professionals and four university faculty. In most cases, they did not have specific knowledge or expertise matching the learning objects' subject matter. The participants were located in Christchurch, New Zealand and Vancouver, Canada. They were offered productivity software in exchange for their participation. The participants were organized into three groups, each with four members.

7. Learning Objects Evaluated

There were eight learning objects selected for evaluation in the study. They were obtained from a variety of sources and included a range of aggregation levels and media. All objects except those originating from the Technical University of British Columbia (TechBC) were obtained through MERLOT. The learning objects were randomly assigned to be discussed in the collaborative session (set A) or not (set B).

8. Research Design and Procedure

In phase one of the experiment, participants were asked to use LORI individually to evaluate the objects in set A and set B in a fixed order without backtracking. This was done using a Microsoft Word document containing hyperlinks to the eight learning objects, with data capture using a Microsoft Excel spreadsheet. These documents were emailed to the participants in advance. Once the evaluation was completed by the participant, the spreadsheet was emailed back to one of the researchers, who assembled the results from all participants into a single spreadsheet and obtained means and standard deviations.

Two days later, in phase two, the teams of four used text chat to discuss the learning objects in set A. During the chat sessions, each participant was able to view an individualized spreadsheet showing the distributions of phase one ratings with only their own ratings identified. The discussion was moderated by one of the researchers to ensure all four objects were discussed. The sessions lasted 1 h with each object receiving about 20 min of discussion. The moderator used the initial analysis of LORI ratings to prioritize the items. Items with the greatest inter-rater variation were discussed before items with lower inter-rater variation. There was usually sufficient time to discuss about four or five of the items. The collaborative assessment sessions were held using the chat tool in the MSN Groups Website, and were scheduled during hours suitable to the geographically dispersed participants (Canada and New Zealand).

In phase three, scheduled on the fifth day, the participants individually re-evaluated sets A and B. They also completed a questionnaire that asked for their views on a range of topics related to the research goals of the study.

This design enabled an investigation of the effects of structured collaborative assessment. Set B served as a control to separate the effects of collaboration specific to the objects discussed from those that generalize to all learning objects.

9. Results

Table 3 shows the inter-rater reliability for the first eight items of LORI. Each cell of the table contains two data elements: the pre-discussion result (first) and the post-discussion result (second) connected by an arrow. Using SPSS, we obtained the intraclass correlation (ICC), two-way random model with absolute agreement [20]. ICCs for items 9 and 10 are not presented because of insufficient variation between learning objects and violation of the assumption of normality. Judging from participant comments, this was caused by their lack of knowledge of the metadata and accessibility standards. The reliabilities in Table 3 were obtained with the eight learning objects serving as separate targets. In constructing this table, two raters were dropped from the analysis due to missing data.

Table 3
Reliability of Eight LORI Items before
and after Discussion

Item	Single ICC	Average ICC (10 Raters)	Alpha (10 Raters)
1	$0.18 \rightarrow 0.28$	$0.69 \rightarrow 0.80$	$0.77 \rightarrow 0.82$
2	$0.22 \rightarrow 0.30$	$0.73 \rightarrow 0.81$	$0.78 \rightarrow 0.83$
3	$0.13 \rightarrow 0.17$	$0.60 \rightarrow 0.67$	$0.72 \rightarrow 0.74$
4	$0.21 \rightarrow 0.19$	$0.73 \rightarrow 0.70$	$0.79 \rightarrow 0.74$
5	$0.43 \rightarrow 0.43$	$0.88 \rightarrow 0.88$	$0.90 \rightarrow 0.90$
6	$0.53 \rightarrow 0.53$	$0.92 \rightarrow 0.92$	$0.93 \rightarrow 0.93$
7	$0.55 \rightarrow 0.75$	$0.92 \rightarrow 0.97$	$0.93 \rightarrow 0.97$
8	$0.13 \rightarrow 0.27$	$0.60 \rightarrow 0.79$	$0.60 \rightarrow 0.84$

The first value in each cell of the single ICC column is the estimated reliability for a single rater, with no training, prior to a group discussion of specific learning objects. As would be expected for a scale comprised of a single item, these values are all well below the threshold of 0.75 or 0.80 that might be regarded as sufficient reliability for this application.

The first value in each cell of the average ICC column is the reliability expected when the pre-discussion ratings of 10 raters with no training are averaged. Under these conditions, three of the eight items exceed the arbitrary threshold of 0.80. Discussion of specific learning objects tended to increase the Average ICC to the extent that six of the eight items show sufficient reliability.

The right-most column of Table 3 shows Cronbach's alpha as an index of inter-rater (not inter-item) consistency. Because alpha is insensitive to rater differences that are linear transformations, it can be compared with Average ICC to detect consistent rater bias. A greater difference between the two coefficients indicates a greater rater bias. The data are consistent with the hypothesis that discussion of specific learning objects tends to reduce rater bias.

Like the analysis of variance, ICC is calculated as a comparison of variation between cases (e.g., learning objects) to variation within cases (e.g., across raters). When the variation across cases is large and variation within cases is small, large positive values approaching 1.0 are obtained.

When there is not sufficient variation between cases, the ICC may be very low and possibly negative, even when there is substantial agreement among raters. This situation existed for items 9 and 10. For example, with all eight learning objects, the value 0 was the modal rating for item 9 (Metadata and Interoperability Compliance). This resulted in ICCs approaching 0 despite the fact of majority agreement among raters as shown in Table 4.

Table 4
Mean Percent Agreement on
Mode for Items 9 and 10

Item	Agreement on Mode (%)
9	$67 \rightarrow 81$
10	$65 \rightarrow 64$

Table 4 was constructed by taking the mean of the percent of ratings at the modal value over eight learning objects. Due to missing data, the number of raters included varied from 8 to 11.

Table 5 shows inter-rater reliabilities for control and experimental (discussed) learning objects before and after collaborative sessions. The reliabilities are Averaged ICC based on 10 raters. Two raters were dropped because of missing data.

Table 5
Control versus Experimental
Learning Objects

Item	Control	Experimental
1	$0.40 \rightarrow 0.20$	$0.71 \rightarrow 0.91$
2	$0.16 \rightarrow 0.46$	$0.78 \rightarrow 0.87$
3	$0.37 \rightarrow 0.47$	$0.60 \rightarrow 0.64$
4	$0.59 \rightarrow 0.57$	$0.57 \rightarrow 0.64$
5	$0.78 \rightarrow 0.76$	$0.90 \rightarrow 0.92$
6	$0.93 \rightarrow 0.92$	$0.80 \rightarrow 0.86$
7	$0.92 \rightarrow 0.94$	$0.91 \rightarrow 0.96$
8	$0.39 \rightarrow 0.73$	$0.76 \rightarrow 0.83$

The last column of Table 5 shows consistent improvement in inter-rater reliabilities as a result of the collaborative discussion of those objects. This is in contrast to the control objects, which were not discussed. The control

objects returned mixed results, with some items showing improved inter-rater reliabilities, while others are worse.

10. Conclusion

This study has provided evidence that LORI can be used to reliably assess some aspects of learning objects and that using a collaborative assessment process can improve inter-rater reliability. It has presented inter-rater reliabilities based on aggregation of the ratings provided by 10 raters. Although this provides a useful estimate of the number of raters currently required to obtain reliable assessment, our goal is to improve the instrument and collaboration process to the point where a single collaborative group with four to six raters is sufficient.

Modifications being considered include:

- Introduce prior training to increase overall reliability. This seems especially crucial with regard to items dealing with compliance to metadata, interoperability, and accessibility specifications.
- Ensure that at least two raters have expertise in the subject matter dealt with by the learning object. This is expected to greatly improve reliability and validity of ratings related to content accuracy.
- Revise rubrics comprising several of the items based on comments received from participants in the study. This is expected to especially increase reliability for items covering content accuracy, support for learning goals, reusability, and compliance to existing specifications.

The LORI instrument can assist designers and developers in summatively evaluating the quality and usefulness of existing learning objects. It can also serve as a formative device for improving learning objects under development. Aside from producing more reliable and valid evaluations, collaborative online evaluation that focuses on discrepancies among ratings holds great promise as a method for building professional communities in relation to learning objects.

Acknowledgements

References

[1] P. Munkittrick, Building a foundation for connected learning, *T.H.E. Journal*, *27*(9), 2000, 54–56.
[2] D. Caterinicchia, NSA invests in e-learning, *Federal Computer Week*, *14*(16), 2000, 18.
[3] Learning Technology Standards Committee (LTSC) of the IEEE. Retrieved August 2002. Available at http://ltsc. ieee.org.
[4] IEEE P1484.12 Learning Object Metadata Working Group. Scope and purpose. Retrieved July 2002. Available at http://ltsc.ieee.org/wg12/s_p.html.
[5] Draft Standard for Learning Object Metadata. IEEE P1484.12.1. Retrieved August 2002. Available at http://ltsc. ieee.org/doc/wg12/LOM_WD6_4.pdf.
[6] IMS Global Learning Consortium. Retrieved August 2002. Available at http://www.imsproject.org.
[7] ARIADNE Foundation. Retrieved August 2002. Available at http://www.ariadne-eu.org.
[8] Dublin Core Metadata Initiative. Retrieved August 2002. Available at http://dublincore.org.
[9] ADLNet. Retrieved August 2002. Available at http://www. adlnet.org.
[10] T. Barron, Learning object pioneers. Retrieved July 2002. Available at http://www.learningcircuits.org/mar2000/ barron.html.
[11] M. Hatala & G. Richards, *Global vs. community metadata standards: Empowering users for knowledge exchange*, ed. I. Horrocks & J. Hendler, ISWC 2002, Springer, LNCS 2342, 2002, 292–306.
[12] D. Wiley, Peer-to-peer and learning objects: The new potential for collaborative constructivist learning online, *Proc. of the International Conf. on Advanced Learning Technologies*, 2001, 494–495.
[13] J.R. Hill & M.J. Hannafin, Teaching and learning in digital environments: The resurgence of resource-based learning, *Educational Technology Research and Development*, *49*(3), 2001, 37–52.
[14] R.A. Reiser & H.W. Kegelmann, Evaluating instructional software: A review and critique of current methods, *Educational Technology Research and Development*, *42*(3), 1994, 63–69.
[15] D.A. Wiley, Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, in D.A. Wiley (Ed.), *The instructional use of learning objects*: Online Version 2000. Retrieved August 2002. Available at http://reusability.org/read/chapters/wiley.doc.
[16] Southern Regional Education Board. *EvaluTech*. Retrieved August 2002. Available at http://www.evalutech.sreb.org.
[17] ASTD draft certification standards. Retrieved July 2002. Available at http://www.astd.org/ecertification.
[18] K. Belfer, J.C. Nesbit, A. Archambault, & J. Vargo, Learning Object Review Instrument (LORI) Version 1.3, 2002. Retrieved August 2002. Available at http://www. sfu.ca/~kbelfer/LORI/lori13.rtf.
[19] J.C. Nesbit, K. Belfer, & J. Vargo, A convergent participation model for evaluation of learning objects, *Canadian Journal of Learning and Technology*, *28*(3), 2002, 105–120.
[20] P.E. Shrout & J.L. Fleiss, Intraclass correlations, *Psychological Bulletin*, *86*(2), 1979, 420–428.

Biographies
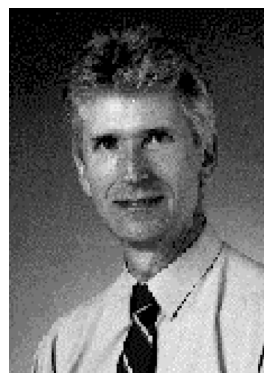
John Vargo is Dean of Commerce and member of the Department of Accountancy Finance and Information Systems at the University of Canterbury in Christchurch, New Zealand. He has published three books, and over 50 research articles in refereed journals, conferences, and other sources. John obtained an MBA from the University of Santa Clara in California and a Ph.D. from the University of Canterbury. His research interests are in e-commerce, strategic use of information systems and the effective use of learning technologies.

John Nesbit is an Associate Professor in the Information Technology and Interactive Arts program at Simon Fraser University in British Columbia, Canada. He has published over 25 refereed articles in journals and conference proceedings. John completed a Ph.D. in Educational Psychology at the University of Alberta in 1988 and undergraduate studies in psychology at the University of British Columbia. John's research interests include evaluation models, eLearning delivery models, cognition and pedagogy in eLearning, virtual communities, and adaptive instructional systems.

Karen Belfer is the Program Evaluation and Assessment Coordinator at the E-Learning and INnovation Center (E-LINC) at Simon Fraser University in British Columbia, Canada. Karen did her undergraduate work in Informatics and her Masters in Education at the Anahuac University in Mexico, where she taught for over 10 years. Karen has extensive experience in the use of technology in higher education. Her professional and research interests are in faculty development and the assessment of online social learning environments, learning objects, and teamwork.

Anne Archambault is a Program Manager at Microsoft Corporation, where she designs online collaboration tools for Microsoft Office. She worked as Educational Multimedia Production Manager for the Technical University of British Columbia and as Product Manager for Brainium.com. In 2000, she received an EDUCAUSE/NLII fellowship that focused on online communities. Anne holds a bachelors degree of Microbiology and Immunology from McGill University and a Masters of Environmental Studies from York University. Her research interests include online collaboration and virtual communities.