Canadian Journal of Learning and Technology
Volume 28(3) Fall / automne, 2002
A Convergent Participation Model for Evaluation of Learning Objects

John Nesbit

Karen Belfer

John Vargo

About the Authors

John Nesbit is an Associate Professor at Simon Fraser University. He is currently conducting research on learning object evaluation, cooperation in online learning environments and scalable models for delivery of eLearning.

Karen Belfer is the Program Evaluation and Assessment Coordinator at the eLearning and INnovation Center (eLINC) at Simon Fraser University. She is currently completing Doctoral research on course climate in online learning with the University Complutense of Madrid.

John Vargo is the Dean of Commerce at the University of Canterbury in Christchurch, New Zealand. His research interests are in e-commerce, strategic use of information systems and the effective use of learning technologies.

Correspondence concerning this article should be addressed to John Nesbit, Simon Fraser University Surrey, 2400 Central City, Surrey BC, V3T 2W1, Canada. Tel: 604-586-5247 Email: nesbit@sfu.ca

Abstract

The properties that distinguish learning objects from other forms of educational software - global accessibility, metadata standards, finer granularity and reusability - have implications for evaluation. This article proposes a convergent participation model for learning object evaluation in which representatives from stakeholder groups (e.g., students, instructors, subject matter experts, instructional designers, and media developers) converge toward more similar descriptions and ratings through a two-stage process supported by online collaboration tools. The article reviews evaluation models that have been applied to educational software and media, considers models for gathering and meta-evaluating individual user reviews that have recently emerged on the Web, and describes the peer review model adopted for the MERLOT repository. The convergent participation model is assessed in relation to other models and with respect to its support for eight goals of learning object evaluation: (1) aid for searching and selecting, (2) guidance for use, (3) formative evaluation, (4) influence on

design practices, (5) professional development and student learning, (6) community building, (7) social recognition, and (8) economic exchange.

Rsum

Les propriétés qui distinguent les objets d'apprentissage des autres formes de logiciels éducatifs : accessibilité mondiale, normes de métadonnées, granularité et réusabilité, ont des implications en termes d'évaluation. Cet article propose un modèle de participation convergente en vue de l'évaluation des objets d'apprentissage dans lequel les représentants des parties prenantes (par exemple, les étudiants, les chargés de cours, les experts en la matière, les concepteurs de matériel pédagogique et les développeurs de médias) convergent vers des descriptions et des notations plus semblables par l'intermédiaire d'un processus à deux étapes assisté par des outils de collaboration en ligne. L'article passe en revue les modèles d'évaluation qui ont été appliqués aux logiciels et aux médias éducatifs, considère certains modèles de regroupement et de méta-évaluation des revues d'utilisateurs individuels, qui sont récemment apparues sur le Web et décrit le modèle de contrôle par les pairs adopté pour la logithèque de données MERLOT. Le modèle de participation convergente est évalué par rapport à d'autres modèles et en termes de soutien pour huit objectifs d'évaluation des objets d'apprentissage : (1) aide pour la recherche et la sélection, (2) guide pour l'utilisation, (3) évaluation formative, (4) influence sur les pratiques de la conception, (5) développement professionnel et apprentissage de l'étudiant, (6) développement d'une conscience communautaire, (7) reconnaissance sociale et (8) échange économique.

Introduction

When instructors and students search a repository to select a learning object for use, they have three questions in mind: Is it the right type? Is it the best I can find? and How should I use it? As is documented in other articles in this special issue, substantial effort has been spent in recent years building metadata standards and tools to assist users in answering the first question. Now, with repositories filling with thousands of objects, many dealing with the same subject matter and learning goals, greater attention and effort will turn to the second and third questions.

Learning object evaluation by third-party reviewers is a key element in promoting reusability because the availability of quality ratings can have an immediate and compelling effect on the outcome of repository searches. Quality ratings are already being used to order search results in MERLOT, North America's leading learning object metadata repository for higher education. In MERLOT, highly rated objects are returned ahead of objects that have lower ratings or have not been evaluated. The use of quality ratings in this way places significant responsibility on the eLearning research community to develop the best possible evaluation methods because even minor variations in the method may have amplified effects on many thousands of users' selection decisions.

The challenge of developing effective evaluation systems is formidable because they must optimize on two opposing variables: the number of objects that can be evaluated versus the quality of the obtained evaluations. A costly model that returns highly accurate evaluations is of little use to the individual user if it can only be applied to a small fraction of a collection. Also, it is insufficient for reviews to return only numeric quality ratings: Reviews can most strongly benefit users by elaborating on pre-existing metadata to describe the situations and contexts in which the objects can be appropriately used.

Some definitions of learning objects admit non-digital resources and cover a wide range of aggregation levels including whole courses and certificate programs (IEEE, 2002). But because our goal is to facilitate reusability through quality evaluation, this paper is primarily concerned with digital resources at the lower to middle levels of granularity; a scope that includes static images or text, animations, simulations, interactive lessons, tests, and so on.

We have identified 8 interrelated reasons for developing effective learning object evaluation systems:

• Ratings and qualitative assessments aid individual users in searching and selecting objects.

• Evaluations can provide guidance on how best to use an object.

• Quality can be increased by formative evaluation throughout the design and development stages.

• Evaluation standards can drive the practices of designers and developers.

• Participation in evaluation activities can contribute to the professional development of those who work with learning objects.

• Evaluation activities can build and support communities of practice in relation to learning objects.

• Positive evaluations can promote social recognition of skilled designers and developers.

• A trusted evaluation system may be an essential step toward the development of a workable business model for the economic exchange of learning objects.

Here we propose a convergent participation model that addresses all of the above goals. The model is defined by the following characteristics:

• Evaluation is conducted by a panel drawn from different stakeholder groups. For example, in a college setting, learners, instructors, instructional designers, and media developers might participate as panel members.

• There is a two-cycle process. During the first cycle the evaluators independently and asynchronously assess the object. During the second cycle, which may be conducted synchronously or asynchronously, the evaluators compare and explain their earlier assessments, adjusting their individual assessments as their judgment shifts in response to the panel's discussion.

• During both cycles the evaluators use an assessment instrument that gathers numeric ratings and verbal comments on a small number of dimensions.

• The process is managed by a moderator who guides discussion but does not directly assess the object. During the second cycle, the moderator sequences the items under discussion to ensure that features about which there is greatest disagreement are discussed first.

• Evaluators use online tools that enable fully distributed participation. The tools support the communication of individual assessments among group members and aggregation of individual assessments into collaborative reviews.

Evaluation of Educational Software and Multimedia

In considering some of the alternative approaches to evaluating educational software and multimedia, we use the classification scheme of Worthen, Sanders and Fitzpatrick (1997) who distinguish between, among others, consumer-oriented, expertise-oriented, objectives-oriented, and participant-oriented evaluations.

Consumer-oriented evaluations of educational materials are conducted by governmental organizations and non-profit associations that train evaluators, often teachers, to apply standard criteria, checklists or rating scales, to examine the materials and produce reviews in a highly structured format. With costs in the same range as those for the production of basic learning object metadata, consumer-oriented evaluation is one of the least expensive models, and for this reason alone is highly applicable to the problem of assessing large numbers of learning objects. Containing over 7000 reviews of K-12 materials, the EvaluTech repository (Southern Regional Education Board) is perhaps the best evidence of the cost efficiency of a consumer-oriented model. The quantitative instruments often used in consumer-oriented evaluation can provide a score for each object that can be used to order search results. Also, structured reporting formats have the advantage of facilitating comparison among objects.

The validity and descriptive power of consumer-oriented evaluations are limited by the same factors that make them cost efficient. Evaluators typically work individually, unable to benefit from the specialized expertise of others. Depth of analysis, and flexibility to deal with special characteristics of objects, are restricted to whatever can be built into the instrument and procedure on which the evaluator is trained. In reviewing instruments used for educational software evaluation, Gibbs, Graves, and Bernas (2001) noted that they "have been criticized for not being comprehensive, understandable, and easy to use." Within criteria that they do attempt to evaluate — usually content, interface design and technical operation — there is evidence of low correlations among evaluators (Jolicoeur & Berger, 1986).

Expertise-oriented evaluations are conducted by recognized experts, either individually or in panels. In comparison with consumer-oriented approaches, they place less emphasis on structured formats, closed-response instruments, and training; relying instead on the expert judgment of the evaluator. The MERLOT peer review process, described in a later section of this article, can be regarded as combining the expertise and consumer-oriented approaches because it brings expert evaluators together with standard scoring criteria and structured reports.

Expertise-oriented approaches have been criticized as especially vulnerable to the subjective biases of the evaluator. And they often show low inter-evaluator consistency because expert evaluators may vary widely in the importance placed on different factors (McDougall & Squires, 1995; Reiser & Dick, 1990). As in the consumer-oriented model, there is no representation of stakeholders such as learners, and no testing of the object in situ . There may be representation from only a single expert community - perhaps only subject matter experts or only instructional design experts. Expertise-oriented approaches tend to be more costly than consumer-oriented approaches because they lack the efficiency advantages of a fixed procedure repeatedly applied by a trained evaluator.

Objectives-oriented approaches couple detailed analysis and definition of goals with empirical, quantitative studies using pre-post or comparative designs that test the extent to which the goals have been attained. Reiser and Dick (1990) developed a model for educational software evaluation that involved, in part: (1) defining learning objectives through an analysis of the software and its documentation, (2) developing test items and attitude questions based on the objectives, (3) conducting an initial trial with three representative students and (4) conducting a similar trial with a larger group of students in the targeted learning environment. Similar

models have been used for demonstrating the efficacy of software for teaching skills such as spelling and fractions (Assink & van der Linden, 1993; Jolicoeur & Berger, 1988).

A major drawback of objectives-oriented approaches is the cost associated with running an empirical study. Given our current systems, it is far more expensive to treat and test a group of students than to have a trained evaluator fill out a form. Another limitation of this approach is that in emphasizing goals and outcomes, objective-oriented approaches tend to ignore the learning processes that occur as students interact with materials.

Unlike the consumer-, expertise-, and objectives-oriented approaches, participant oriented approaches to educational evaluation explicitly acknowledge that learning is a social process dependent on social context. This implies that multimedia and software evaluation must account for interactions between the learner and those around them (Baumgartner & Payr, 1996). The methodology of participant-oriented evaluation is drawn from naturalistic and ethnographic research: Data is gathered by prolonged and persistent observation, informal interviews, and document analysis, and reported as detailed descriptions or direct quotations (Neuman, 1989; Patton, 1980). There is often an emphasis on bringing stakeholders together in moderated discussion groups:

Participants in the group process become sensitized to the multiple perspectives that exist around any program. They are exposed to divergent views, multiple possibilities, and competing values. Their view is broadened, and they are exposed to the varying agendas of people with different stakes in the evaluation. This increases the possibility of conducting an evaluation that is responsive to different needs, interests, and values.

Patton (1982, p. 65)

The recognition that program evaluation reports are often underutilized or ignored has led to an interest in promoting organizational learning through the participant group process. Working within an activity theory framework in a university environment, Dobson and McCracken (2001) credited the success of an evaluation project that improved teaching and learning to a team-based approach in which academics and learning technology experts worked as equal partners in furthering innovation. Although the analysis is outside the scope of this paper, activity theory (Nardi, 1996), with its depiction of collective activity (e.g., collaborative evaluation) operating on an object to produce an outcome (e.g., a published review) through the use of mediating tools (e.g., an online evaluation instrument) offers a relevant perspective on all the evaluation approaches we describe, and it is especially consistent with the convergent participation model we are proposing.

Williams (2000) proposed two participant-oriented models that an organization can use to evaluate learning objects it creates. One model deals with externally contracted evaluations of larger instructional units that include learning objects. It involves competitive requests for proposals, stakeholder interviews, possible use of the delphi technique to develop key questions to be addressed by the evaluation, then formulation and execution of an evaluation plan. The second model deals with formative, internally conducted evaluations. It builds an evaluative component into every step of the ADDIE instructional design model (Assess needs, Design, Develop, Implement and Evaluate instruction). Both the internal and external models require meta-evaluation as a final stage.

There are features of most participant-oriented approaches that are incompatible with the requirements of learning object evaluation. Often these approaches tailor the evaluation's guiding questions to situational demands to obtain evaluative reports with varying formats, a practice that would make comparison of large

numbers of learning objects difficult. In participant-oriented approaches there is usually no quantitative rating that can be used to sort search results, an extremely convenient technique that many users have already come to expect. Finally, some other features of participant-oriented approaches are just too costly for use with large numbers of objects.

In the search for an efficient model that preserves many of the advantages of participant-oriented approaches, we were impressed with many of the interactive tools for communication and collaboration provided in online communities. By supporting voluntary contributions from users with automated functions, these websites demonstrate the strengths of technology-mediated communities managed by their members.

Lessons From the Web

User ratings and comments of varying quality have become much easier to gather with the establishment of the Internet. Increasingly sophisticated systems are emerging that address the quality problem by classifying, filtering, and sorting user comments on the basis of ratings provided by other users. The Slashdot website (www.slashdot.org), featuring brief articles that are discussed in hundreds of comments posted by users, led this trend by allowing certain community members to rate others' comments so that any reader could view only comments whose average rating fell above an adjustable threshold. By authoring highly rated comments, a member can acquire the highly sought after "karma points" that increase one's likelihood of being temporarily granted the privilege of rating others' comments. The whole system is extremely effective in ratcheting up the quality of user comments and bringing the most valued comments into the foreground.

The evaluation requirements of a learning object repository are similar in several ways to those of an online book retailer. Amazon's (www.amazon.com) user review and recommendation system offers significant insights into how object repositories might manage user reviews. Any Amazon user can submit a 1000 word book review including a rating on a 5-point scale. Other users can vote yes or no on the usefulness of the review, generating an approval rating that determines the order in which reviews are displayed. The user ratings of a book are combined with a customer's buying preferences and expressed interests to construct a personalized list of recommended books that are presented to the customer when entering the site.

While user reviews offer huge cost advantages over other evaluation models and are, we believe, an essential component of any comprehensive system of learning object evaluation, meta-evaluation mechanisms like those that control review quality on Slashdot and Amazon are more effective on high traffic sites, and for objects that attract many reviewers and meta-evaluators. Perhaps a key to designing whole evaluation systems is to find ways for user reviews and formal reviews to inter-operate in a complementary fashion. Formal reviews by expert or participant panels might bring high quality objects to the attention of more users, or add a critical voice to temper the enthusiastic user reviews of a highly popular object.

Merlot

In contrast to other large metadata repositories, MERLOT (www.merlot.org) has established a set of relatively mature evaluation practices and tools. Emphasizing disciplinary expertise, the approach to learning object evaluation in MERLOT is largely modeled on the academic peer review process for scholarly research and publication familiar to university faculty. The MERLOT web site currently supports 13 discipline-specific communities (e.g., biology, history, information technology, teacher education) and two special interest

communities for faculty development and academic computing professionals. Each discipline community has an editorial board that guides peer review policies and practices for that discipline.

In MERLOT, learning object evaluation is based on 3 dimensions: Quality of Content, Ease of Use, and Potential Effectiveness as a Teaching Tool. Quality of Content encompasses both the educational significance of the content and its accuracy or validity. Ease of Use encompasses usability for first-time users, aesthetic value, and provision of feedback to user responses.

The MERLOT site describes Potential Effectiveness as a Teaching Tool as the most difficult dimension to assess. This dimension encompasses pedagogically appropriate use of media and interactivity, and clarity of learning goals. There is an emphasis on the importance of context, which is defined in terms of learning outcomes, characteristics of the learner, and placement of the materials within a learning strategy. In its guidelines for reviewers, MERLOT makes definite claims regarding the problem of evaluating potentially re-purposable objects:

Remember that the value of "modules" is their ability to be re-purposed for others [to] use in different contexts. In evaluating the Potential Effectiveness for Teaching and Learning, it is CRITICAL to define the purpose of the learning materials. That is, you must contextualize your reviews.

<div align="right">Evaluation Standards for MERLOT</div>

In other words, MERLOT takes the position that it is important for both authors and reviewers to be clear about learning goals. Authors should be clear about what learning goals their materials were designed to support and, because objects may be re-purposed to serve other goals, reviewers should be clear about what learning goals they had in mind during assessment. On this basis we should not be surprised to see different discipline communities offering widely varying reviews on the same object, especially because each discipline community is encouraged to adapt the MERLOT evaluation guidelines to suit its needs.

However, as properties of an object, alignment of learning goals and re-purposability are inherently incompatible. The more an author builds a particular context into a learning object, the less use can be made of it for alternative purposes. One way that purpose is often built into learning materials is to preface them with learning goals or objectives intended to be understood by the learner. It would normally be inappropriate to provide such materials to learners when the goals presented in the materials differ from the learning goals of the situational context. It is notable that, while the MERLOT assessment guidelines mention re-purposing of objects, nowhere do they establish re-purposability as a criterion for evaluation.

A rating scale from 1 to 5 is available and used by over 90% of peer reviews and individual member evaluations. The accompanying rubrics are intended to be applicable in all three dimensions of evaluation (MERLOT):

- Materials not worth using at all.

- Materials do not meet minimal standards but there might be some limited value.

- Materials meet or exceed standards but there are some significant concerns.

- Materials are very good overall but there are a few minor concerns.

- Materials are excellent all around.

By self-enrolling as faculty, staff, student or other, anyone can become a member of MERLOT and submit individual "member comments" on materials registered with the repository. Member comments include an overall rating on the 5-point scale with text fields for general remarks and technical remarks. In addition to applying the 3 dimensions discussed previously, members are asked to indicate the way they evaluated the materials and how much time they spent doing so.

Peer Reviews are usually carried out by two university faculty with relevant content expertise. A third reviewer may be assigned if the two resulting reviews show "significant disparity." For each of the 3 dimensions the reviewers independently and asynchronously assess strengths and concerns, and provide a rating on the 5-point scale. In what might be regarded as a second cycle in the process, one of the reviewers combines the individual assessments and averages the ratings to create an integrated review that is then edited by the other reviewer and discipline co-leaders.

The discipline community leaders tend to select higher quality objects for review (see Table 1). Moreover, the MERLOT website does not report peer reviews with overall ratings of 1 or 2. These practices are intended to encourage developers to register objects in an early stage of development and at the same time highlight exemplary objects.

Of the 6662 objects represented in the MERLOT website at the time of this writing, 21% had at least one member comment and peer reviews were shown for 7%. Table 1 indicates that, for both member comments and peer reviews, a large majority of evaluated objects were rated at level 4 or 5. Only a small proportion of evaluated objects scored below the rating of 3.0 that the rubrics seem to indicate as the cut-off score for acceptability[1] . We conjecture that large sampling bias in favor of higher quality objects will be found in most systems of learning object evaluation, not only MERLOT. Although effort might be invested in culling obsolete or low quality objects on the basis of prima facie judgment, a practice that has been adopted within some MERLOT discipline communities, there is little advantage in allocating significant resources to evaluating low quality items.

Table 1. Ratings from member comments and peer reviews on MERLOT (2002 April 23).

Rating	Member Comments	Peer Reviews
None	126 (9%)	7 (2%)
1.0 – 1.9	41 (3%)	9 (2%)
2.0 – 2.9	88 (6%)	
3.0 – 3.9	302 (21%)	78 (18%)
4.0 - 4.9	541 (38%)	247 (56%)
5.0	319 (23%)	97 (22%)
Total	1417 (100%)	438 (100%)

The MERLOT peer review system was designed to encourage adoption within the academic culture of university faculty. The MERLOT organization's careful attention to the values of that culture, such as professional volunteerism and respect for disciplinary knowledge, seem to have worked well in sustaining growth of the collection, the membership, and the discipline communities. However, in our view, the MERLOT peer review process needs to be extended to match the conditions under which learning objects are developed and used, and to better meet the needs of students. Unlike journal articles written by and for academic researchers, learning objects are created by teams of various types consisting of faculty, instructional designers, media developers and programmers; and they are used by instructors and students. As we have

discussed earlier, assessment models that include representative participation from all stakeholder groups are far more likely to result in valid evaluations.

Convergent Participation

The convergent participation model, depicted in Figure 1, is a panel evaluation process designed to obtain better outcomes than the peer review model without resorting to expensive field studies. When applied to the evaluation of learning objects, it is presumed to exist in an online community, similar to MERLOT, in which users post individual member reviews that include comments and ratings. Panel reviews are organized by a moderator who selects one or more objects for review according to criteria established by the community. The moderator also selects and invites panel participants that represent different stakeholder groups, some of whom may have already posted individual member reviews of the object(s) to be evaluated by the panel.
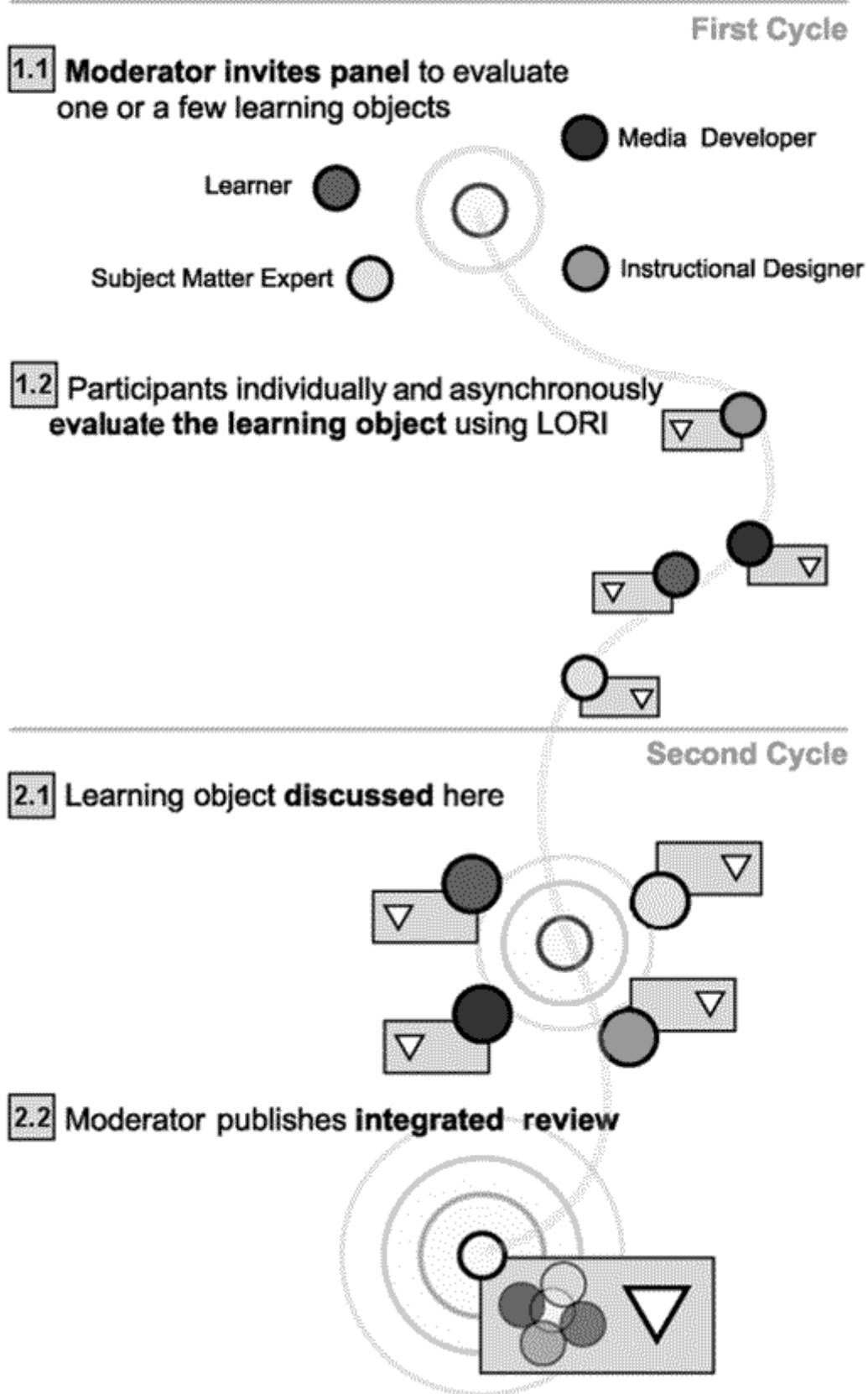
Learning object published

1.1 **Moderator invites panel** to evaluate one or a few learning objects

Media Developer

Learner

Subject Matter Expert

Instructional Designer

1.2 Participants individually and asynchronously **evaluate the learning object** using LORI

2.1 Learning object **discussed** here

2.2 Moderator publishes **integrated review**

Figure 1. Convergent Participation with LORI

In both cycles of this two-cycle model, the reviewers use an evaluation instrument capable of gathering ratings and comments specific to several different features of learning objects. In the first cycle, through an asynchronous process lasting several days, the participants examine the object and submit an individual review

no different from those they might post outside the panel evaluation process. Those who had already submitted a review might edit their existing review or simply revisit it to recall the reasoning behind their judgment.

In the second cycle, the reviews from all participants are exposed in an integrated format, and the moderator leads a discussion focusing on the points of greatest divergence among the participants. The moderator's role is to keep the discussion on track without revealing views he or she may have about the object or otherwise biasing the judgment of the participants. A critical characteristic of the model is that the reviews produced in the first cycle are used by the moderator to sequence the discourse so that features of the learning object about which there is least agreement are discussed first. As the second cycle proceeds the participants may edit their individual reviews using tools that immediately update the integrated view available to the entire panel. When the time allocated for the second cycle expires, the moderator brings the discussion to a close and asks all participants to approve the publication of the panel review consisting of the integrated ratings and comments. Integrated reviews show the range and central tendency of individual ratings with comments concatenated within different evaluative categories. Only data from participants who approve publication are included in published review.

Several features of the model have been left undetermined: some because they seem situation dependent, and others because we have insufficient experience with the model to make a well-founded recommendation. Our preference is to conduct the second cycle as a real-time meeting supported by synchronous communication tools, but there may be conditions under which asynchronous discussion would be more suitable. Anonymity of the participants, both among themselves and in relation to the larger community, is a potentially important element that invites further investigation.

In a previous study we tested an early prototype of the convergent participation model with 12 participants divided into three internationally distributed panels (Vargo et al., in press). The participants included instructional designers, university faculty and media developers. Each participant evaluated 8 learning objects: 4 objects were individually evaluated and 4 were evaluated through convergent participation. The participants used the Learning Object Rating Instrument (LORI) consisting of the items shown in Table 2. LORI presents a 5-point scale for each of its 10 items. Inter-rater reliability was measured by intraclass correlation (Shrout & Fleiss, 1979). Inter-rater reliability was measured for first-cycle and second cycle evaluations. The results showed more consistent convergence across all raters and panels (i.e., increased inter-rater reliability) for the collaboratively evaluated objects than the individually evaluated objects. Some participants spontaneously commented that the process was an excellent way to develop their knowledge about learning objects. Participant comments also established that training on some of the dimensions of LORI is a requirement for effective use of the instrument.

Table 2. Items in LORI (version 1.3)

1. **Presentation Aesthetics**
   Selection of fonts and colors, writing style, legibility of text, and other production values

2. **Presentation Design for Learning**
   Formatting, highlighting and sequencing of information

3. **Accuracy of Content**
   Veracity, accuracy, level of detail, and educational significance of knowledge

4. **Support for Learning Goals**
   The alignment of learning activities, content and assessment to goals

5. **Motivation**
   Ability to motivate, and stimulate the interest of, an identified population of learners

6. **Interaction Usability**
   Ease of navigation, predictability of the user interface, and the quality of UI help features

7. **Interaction Feedback and Adaptation**
   Adaptive content driven by differential learner input or learner modeling

8. **Reusability**
   Ability to port between different courses or learning contexts without modification

9. **Standards Compliance**
   Adherence to international standards and specifications

10. **Accessibility**
    Support for learners from diverse backgrounds and with varying capabilities

There are points of similarity between the Convergent Participation model and the practice of focus groups as described by Kreuger (1994). As with focus groups, the moderator's role is to facilitate a conversation among the participants, not to conduct a series of interviews with individual participants. In addition to pacing the discussion and maintaining focus, the moderator must establish an environment where participants feel comfortable to express differing views and support their views through argumentation.

Although there is evidence that participants, in both the proposed model and in focus groups (Kreuger, 1994), do converge in their views as a result of discussion, the emergence of consensus is not a required or even desired outcome of either approach. In using the convergent participation model, one hopes that participants will come to a common understanding of the instrument and how it applies to the learning object, but it is expected that reviews will often be published that show disagreement among the reviewers.

One might question why the whole panel is expected to judge dimensions on which only some participants have professional expertise. Although in principle participants can select a don't know option, the moderator does encourage all participants to contribute a rating for each dimension. This practice invites any expert on the dimension under consideration to explain the rationale behind his or her ratings and comments so that others are persuaded to rate in a similar way. When two experts disagree, the other ratings will sway to the side that can most effectively communicate its position to the `lay' participants.

Support for the Goals of Learning Object Evaluation

In the introduction to this article we offered the claim that the convergent participation model addresses 8 goals for learning object evaluation. We are now in a position to discuss the extent of support it provides for each of these goals, relative to support provided by other models.

• Aid for searching and selecting. Like the peer review model, the convergent participation model is likely too costly to fully cover a large repository containing many thousands of learning objects. We believe that it can complement individual user reviews and consumer-oriented reviews by offering better quality evaluations in high demand areas and for objects that for any reason require greater attention and analysis. We hope that, by drawing a broader cross-section of participants into the collaborative review process, a larger base of trained reviewers can be established to carry out high quality individual reviews. The model seems well suited to comparative reviews that evaluate a few objects occupying overlapping curricular space, although it has yet to be extended or tested for that application.

• Guidance for use. The extent to which a convergent participation review suggests ideas to instructors and learners about how best to use an object depends on the nature of the instrument that structures the discussion. We would argue, though, that a discussion on lesson planning or assignment design that includes instructional designers, learners, subject matter experts and instructors will offer greater insight than one in which only one of these roles is represented.

• Formative evaluation. Although convergent participation reviews can certainly reveal strengths and weaknesses in an early version of an object in a format that would be very useful for a development team, in its current form the model does not offer the continuous monitoring that is regarded as optimal for formative evaluation. The model could be quite usefully extended to work within an instructional design or project planning model, perhaps by substituting different instruments as appropriate for use at different project milestones. With regard to points (2) and (3) we caution that, as an evaluation model, convergent participation cannot substitute for the range of functions carried out by an instructional designer.

• Influence on design practices. Regardless of whether one is considering curricular, instructional, or media design, an evaluation standard will tend to drive design practices if it represents the values of the community of designers and developers. When one community determines the standards used to evaluate the work of other communities, tensions arise between communities, evaluations are underutilized, and individual design practices are unaffected. Convergent participation is a boundary crossing activity that should assist in establishing common evaluation standards that are accepted across represented communities and are capable of influencing individual practice within each community.

• Professional development and student learning. When communicating across the boundaries of disciplinary expertise it is often necessary to expose tacit assumptions and press implicit knowledge into explicit language. In our view this has great educative value for all concerned and forms ideal conditions for professional development. Further, if the instrument adopted with the model is derived from research, the participants are naturally led into an increased familiarity and critical understanding of the underlying research base. Under some conditions, it seems justifiable to assign course credit or professional development release time for work on convergent participation review panels.

• Community building. The work of evaluating learning objects in an intensely collaborative setting both expends, and creates in greater measure, what Putnam (2000) calls social capital, the pro-social motive normally generated in any matrix of active interpersonal relationships. Social capital is community-building

glue that can drive community members to act in the common interest, perhaps contributing unpaid hours to observe learners interacting with a new user interface, or spending a portion of sabbatical time on re-designing a curriculum. Because the activity of developing high quality learning objects requires the cooperation of different professional groups, it is important that social capital generated in the evaluation activity be spread across those groups to form an inclusive community.

• Social recognition. Although many different evaluation systems can inform the granting of professional awards, these surely carry greater social recognition value when established in a social context wider than the specialized professional guild to which the recipient belongs. A media development award will garner more respect among professors when professors were represented in the evaluative process that granted it.

• Evaluation for economic exchange. The quality of evaluative information is very often a deciding factor in the purchase of any commodity. The availability of higher quality evaluations may reduce licensing costs by eliminating the need for the purchaser to carry out in-house evaluations. We argue that evaluation from any single disciplinary perspective, whether that is content, instructional design, or media design, is insufficient; and further that those who buy learning objects to assemble commercially offered courses are most interested in the views of the end-user, the learner.

To summarize, the strengths of the convergent participation model are that it brings together representatives of stakeholder groups, efficiently focuses their attention on the points that may be in greatest need of resolution, and produces a review that concisely presents areas of agreement and dissent among the evaluators. Advancement of the model requires more study of how converging participants interact, and development of tools to support that interaction.

Acknowledgments

References

Amazon. Retrieved April 25, 2002 from http://www.amazon.com

Assink, E., & van der Linden, J. (1993). Computer controlled spelling instruction: A case study in courseware design.

Journal of Educational Computing Research

,

9

(1), 17-28.

Baumgartner, P., & Payr, S. (1996). Learning as action: A social science approach to the evaluation of interactive media. Annual Conference on Educational Multimedia and Hypermedia. Association for the

Advancement of Computing in Education. Retrieved April 27, 2002, from
http://www.webcom.com/journal/baumgart.html

Dobson, M., & McCracken, J. (2001). Evaluating technology-supported teaching and learning: A catalyst to
organizational change. Interactive Learning Environments, 9(2), 143-170.

Gibbs, W., Graves, P. R., & Bernas, R. S. (2001). Evaluation guidelines for multimedia courseware. Journal of
Research on Technology in Education , 34(1), 2-17.

IEEE Learning Technology Standards Committee (2002). Learning Object Metadata Standard IEEE 1484.12.1.
Retrieved September 7, 2002, from http://ltsc.ieee.org/wg12/

Jolicoeur, K., & Berger, D. E. (1986). Do we really know what makes educational software effective? A call
for empirical research on effectiveness. Educational Technology,26(12), 7-11.

Jolicoeur, K., & Berger, D. E. (1988). Implementing educational software and evaluating its academic
effectiveness: Part I. Educational Technology, 28(9), 7-13.

Krueger, R. A. (1994). Focus groups: A practical guide for applied research (2nd edition). London: Sage.

McDougall, A., & Squires, D. (1995). A critical examination of the checklist approach in software selection.
Journal of Educational Computing Research, 12(3), 263-274.

MERLOT: Multimedia Educational Resources for Learning and Online Teaching. Retrieved April 17, 2002,
from http://merlot.org/

Nardi, B. (1996). Context and consciousness: Activity theory and human-computer interaction. Cambridge,
MA: MIT Press.

Neuman, D. (1989). Naturalistic inquiry and computer-based instruction: Rationale, procedures, and potential.
Educational Technology, Research & Development, 37(3), 39-51.

Patton, M. Q. (1980). Qualitative evaluation methods. Beverly Hills: Sage.

Patton, M. Q. (1982). Practical evaluation. Beverly Hills: Sage.

Putnam, R. (2000). Bowling alone: The collapse and revival of American community [Electronic version].
New York: Simon & Schuster.

Reiser, R. A., & Dick, W. (1990). Evaluating instructional software. Educational Technology, Research and
Development, 38 (3), 43-50.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations. Psychological Bulletin, 86, 420-428.

Slashdot. Retrieved April 23, 2002, from http://www.slashdot.org/

Southern Regional Education Board. EvaluTech. Retrieved April 24, 2002, from
http://www.evalutech.sreb.org/

Vargo, J., Nesbit, J., Belfer, K., & Archambault, A. (in press). Learning object evaluation: Computer mediated
collaboration and inter-rater reliability. International Journal of Computers and Applications.

Williams, D. D. (2000). Evaluation of learning objects and instruction using learning objects. In D. A. Wiley
(Ed.), The instructional use of learning objects [Electronic version]. Retrieved April 18, 2002 from

http://reusability.org/read/chapters/williams.doc

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). Program evaluation: Alternative approaches and practical guidelines (2nd ed.). New York: Longman.

Endnotes

1. We can distinguish between pre-review and post-review sampling bias. In MERLOT, learning object developers can request that reviews not be posted or that their objects be de-registered from the repository.

———————————————