A brief overview of quality inside learning object repositories

Cristian Cecchinel  
Federal University of Pelotas (UFPEl)  
R. Felix da Cunha 630, Centro – Pelotas (RS), Brasil  
+55 (53) 32279079  
contato@cristiancechinel.pro.br

Xavier Ochoa  
Escuela Superior Politécnica del Litoral (ESPOL)  
Km. 30.5 Vía Perimetral- Guayaquil, Ecuador  
+593 4 2269773 ext. 7006  
xavier@cti.espol.edu.ec

ABSTRACT
Assessing quality of learning objects (LOs) is a difficult and complex task that normally revolves around multiple and different aspects that need to be addressed. Nowadays, quality assessment of LOs inside repositories is based on the information provided by the same community of users and experts that use such platforms. These kinds of information are known as evaluative metadata and constitute a value body of knowledge about LOs that is normally used inside the repositories during the process of searching and retrieving. The present work aims to present a brief overview about how LOs quality is being assessed inside some of the most important repositories existent nowadays, as well as some limitations of these existing approaches.

Categories and Subject Descriptors
H.3.7 [Information Storage and Retrieval]: Digital Libraries – collection, user issues.

General Terms

Keywords

1. INTRODUCTION
Assessing quality of learning resources is a difficult and complex task that often revolves around multiple and different aspects that must be observed. In fact, the very definition of quality is not straightforward. Vargo, Neshit, Belfer & Archambault [1] state that, even though LO evaluation can be considered a relatively new field, it has roots with an extensive body of prior work on the evaluation of instructional software. As stated by Bethard, Wetzer, Butcher, Martin & Sumner [2] quality is contextual and it will depend on “the alignment between the user constituency being served, the educational setting where deployed, and the intended purpose of the resource”. Vuorikari et al. [3] highlights that existing evaluation approaches could be differentiated based on the process they focus. Among others, they mentioned two characteristic examples of approaches, those which focus on the process of creating resources, and those who focus on ready resources and their evaluation.

According to Williams[4], what a LO ought to be is related to the perspectives of different opinions of those who are the actual users of the resource. So, in order to evaluate quality, it is necessary to consider the particular spectrum of users and the particular set of criteria used by these users to value the resource. Williams [4] proposes a participant-oriented model (involving different users and stakeholders) composed by four types of LO evaluation that should be made simultaneously, repeatedly and sequentially during various stages of the LO development. This approach covers the whole process of creating resources, and the four types of LO evaluation proposed are:

1. Context Evaluation – It tries to establish if there is a need of some LO according to the needs and expectations of the possible users of this LO;
2. Input Evaluation – It compares alternative inputs focusing to meet the needs identified in the previous step. The main goal here is to evaluate the alternative learning objects that could attend the established needs.
3. Process Evaluation – It assesses the planning, the design and the development of the selected inputs, e.g., how well the instructional strategy and LO were implemented.
4. Product Evaluation – It assesses if the LO is attending the initial outcomes expected for its usage.

Each type of evaluation should consider who are the people which care about the LO (the audience of the LO), and what do they care or have interest about. The people who care about the LO could be, for instance, students, teachers, instructional designers, an organization, among others. These audiences can have different understandings and expectations about the LO, and thus can use distinct criteria and values to judge the quality of the LO (for instance, reusability, quality of the metadata, the instructional approach, among others). According to [4], the combination of these information would then define how one should conduct the process of evaluation of a LO.

Besides Williams[4], other authors have also claimed that concerns about quality normally focus on different criteria. For instance, in the context of digital libraries, Custard & Sumner [5] stated that the main issues related to quality are: Accuracy of content, Appropriateness to Intended Audience, Effective Design, and Completeness of Metadata Documentation. In the specific field of learning multimedia resources, the so far most recognized instrument for quantitatively measuring quality is the Learning Object Review Instrument (LORI) [6]. This instrument
is intended to evaluate the final and “ready for use” LO. In LORI quality is evaluated according to nine different criteria which are rated in a 1 to 5 Likert scale (see Figure 1).

Leacock & Nesbit [7] provide some explanations about each one of the nine dimensions of LORI and how they should be interpreted to evaluate LOs:

1. **Content quality** – one of the most important aspects of LO quality. This dimension deals with the level of accuracy and reliability of the content, as well as the existence of biases, errors and omissions.

2. **Learning goal alignment** – it is focused for LOs with a moderate level of granularity, and containing a combination of content, learning activities, and assessments. It intends to evaluate whether the learning activities are aligned with the goals of the LO, and if these activities provide the required knowledge for the users successfully answer the assessments.

3. **Feedback and adaptation** – it measures the capability of the LO to provide feedback and adapt itself according to the user needs. Such adaptation can be related to the localization of the LO for a specific culture or language (as in [8]), or even to change the LO presentation and content according to a certain preferred user learning style, for instance.

4. **Motivation** – it evaluates the ability of the LO in retaining users attention, i.e. if the LO is relevant to the learners’ goals and in accordance to their level of knowledge. According to Leacock & Nesbit [7] learner’s expectations about their success or failure on performing a given task using the LO will also impact on motivation.

5. **Presentation design** – This refers to the quality of exposition (clearness and conciseness) of all items in a LO (text, video, animations, graphics). Aspects such as the font size, or the existence of distracting colors should also be taken into consideration.

6. **Interaction usability** – this criteria evaluates how easy is for a learner to navigate the LO. Good usability will present consistent layout and structure thus avoiding overloading the user with misleading responses and information. Problems with navigation could also be caused, for instance, by broken links or long delays during the usage.

7. **Accessibility** – it refers to accommodation of issues of accessibility of people with disabilities in the design of the LO. For instance, a LO with only textual information would exclude blind learners if no audio voice-over is included.

8. **Reusability** – It deals with the potential of the LO to be used in different courses and contexts. Issues as the granularity of the LO and openness will influence its portability to different scenarios.

9. **Standards compliance** - Whether the metadata fields associated to the LO follow the international standards and are complete in a way that allow others to effectively use that information to search and evaluate the LO relevance.

Even though Leacock & Nesbit [7] provide structural and theoretical foundations for assessing and understanding these many aspects involving quality, they still are all broadly interpreted dimensions that can be subject of divergence from different evaluators. Moreover, different evaluators can also give more importance to one specific dimension than to the others. In order to soften this situation, Nesbit et al. [9] propose applying LORI through the use of a convergent model, where several evaluators from distinct areas (instructors, instructional designers, and multimedia developers) collaborate to achieve a single and unique quality rating for a given resource. In fact, this concept was applied in eLera as it will be shown in next section.

The focus of this paper is to present how evaluation of resources takes place inside learning object repositories. As resources inside repositories are normally ready for use, the quality evaluation approaches adopted by the repositories and covered in the next section are related to the second approach mentioned by Vuorikari et al. [3] (which focus on ready resources rather than on the process of creating them), and to the Product Evaluation type proposed by Williams [4].

The rest of the paper is structured as follows. Section 2 describes how quality assessment takes place inside some of the most important repositories existing nowadays, and presents some basic differences between the two forms of reviews used inside repositories (peer-reviews and public-review). Section 3 discusses some limitations of the current approaches for quality assessment and describes initial results on experiments towards automated quality assessment of learning objects. Section 4 presents the final remarks.

### 2. Evaluation inside Repositories

After their production, LOs must be published in a place where users can easily search and retrieve them for future use, a phase defined in the LO life-cycle by Collis & Strijker [10] as **offering**. Learning Object Repositories (LORs) are the software systems that provide the functionalities for that. A repository could be simply defined as a digital collection where resources are stored for further retrieval. LORs are potential aggregators of communities of practitioners [11-13], i.e. people who share interests and concerns about something they do and learn through their interactions. Due to that, they tend to harness the features of such social environments through the adoption of strategies for the establishment of quality that rely on the impressions of usage and evaluations given by regular users and experts that are members of the repository community. These strategies rely: on 1) the hypothesis of transactive memory systems [14], i.e., systems that store individuals memories, impressions or information about a subject in order to form a universal and collective body of knowledge that can serve as an external memory aid for other individuals; and 2) on the value of metadata from the perspective of social capital theories, i.e.,
enabling and strengthening social relations that have potential to facilitate the accrual of economic or non-economic benefits to the individuals [15].

Vuorikari et al. [3] address this kind of information as evaluative metadata. According to the authors, “evaluative metadata has a cumulative nature, meaning that annotations from different users accumulate by the time, as opposed to having one single authoritative evaluation”. Inside repositories, evaluative information are normally used as the basis for quality assurance of the resources, but also for properly rank and recommend them for users.

In this section we present how evaluative metadata can be found in some of the most important LORs existing nowadays.

2.1 Lera

eLERA (www.elera.net) – Stands for E-Learning Research and Assessment Network. It was a small LOR 1 (with approximately three hundred resources), however, its importance rested on the fact that it was originally created for research purposes. The main focus of the repository was to provide mechanisms and tools for the collaborative and participative assessment of learning objects through the use of LORI.

In eLera, members could create reviews of learning objects by using LORI, and experienced members could moderate teams of members in a collaborative online review process where reviewers discussed and compared their evaluations [15] (see Figure 2). Besides, members could also add some resource to their personal bookmarks, allowing eLera to recommend materials not only by using their associated ratings, but also by using their popularity.

![Figure 2. An eLera request for review (left) and distribution of ratings on a LORI item (right), taken from Nesbit & Li [16]](image)

2.2 Connexions

Connexions is a repository that allows users to collaborative create and share learning materials and that has presented an exponential growth of contributors in the last years. According to [17], such success can be attributed to the fact that, differently from the traditional LORs, Connexions functions through the “social interaction for the creation of materials”, where all materials are created by its own community of members. This community can develop materials in two formats: modules (small pieces of knowledge) and collections (groups of modules structured into course notes). In Connexions every material available is free for using, reusing and sharing with others under a Creative Commons 2 license.

Quality in Connexions is approached by a system called evaluations (see Figure 3) that arranges resources according to evaluations provided by individuals and organizations [18]. In this context, resources are explicitly endorsed by third parties, and gain higher quality assurance as they start to accumulate more endorsements (lenses) from others. Moreover, Connexions also provides mechanisms to sort materials considering their number of accesses over the time and considering the ratings given by users. Recently, Connexions has also integrated to the repository plugins of two popular and well succeeded tools for social interaction (Facebook and Twitter) thus allowing the community of users to recommend and disseminate materials across these social platforms.

2.3 Organic.Edunet

Organic.Edunet (portal.organic-edunet.eu) is a federation of repositories funded by the European Union and focused on contents exclusively related to Organic Agriculture and Agroecology. Even though it is a very recent repository (launched in 2009) it has already approximately 2,500 users and 11,000 resources. The importance of Organic.Edunet also lays on the fact that this repository is a SLOR thus allowing users to perform a semantic search for the materials.

In Organic.Edunet, quality is assured by the community of users who are allowed to rate the resources, and to improve their metadata translations (the portal is multilingual, the interface is available in nine (9) languages and the metadata about the resources is manually translated in up to eight (8) languages)(see Figure 4). Moreover, any user can give direct feedback about a given resource to the portal as well as to report inappropriate contents.

1 Repository that stores only metadata about LOs and not LOs themselves.

2 http://creativecommons.org/
Graphite is a relatively new repository that stores information about learning resources and is supported by Common Sense Media. In Graphite it is possible to find websites, games, and apps that are officially rated by a board of editors and reviewers of the portal. As the portal is built by teachers and for teachers, such community is also allowed to rate and comment the resources, adding impressions of their usage in the classroom (the so called field notes). The averages teachers ratings are then displayed together with the official ratings (see Figure 5). The evaluations range from 1 to 5 and indicate the learning potential of the resources (not for learning, fair, good, very good, excellent) following three learning dimensions, which are:

1. Engagement (whether the resources hold learners’ interest);
2. Pedagogy (if the product carry content central to the learning experience)
3. Supports (whether the resource provides appropriate feedback, and are there support for teachers and learners).

Resources in Graphite are classified/tagged according their subjects (Language and Reading, Math, Science, Social Studies, Arts, and Hobbies) and that the resource facilitates (Thinking and reasoning, creativity, self-direction, emotional development, communication, collaboration, responsibility and ethics, technical skills, and health and fitness). Each resource review also contains comments about the pros and cons of the resource, and how the resource works.

2.5 MERLOT

The Multimedia Educational Resource for Learning and Online Teaching (MERLOT) is a well known and recognized international initiative which allows users to catalogue educational resources aiming to facilitate the use and sharing of online learning technologies[19]. It is developed by the California State University Center for Distributed Learning and stores metadata of over 30,000 materials distributed in several areas (Arts, Business, and Humanities, among others). Its community of users is formed by about 100,000 members. As MERLOT does not store LOs locally, it can be considered as a referatory. The MERLOT repository introduced a post-publication peer-review model in order to assure the quality of its catalogued resources [19]. The materials catalogued in MERLOT are peer-reviewed by different experts in the discipline domain according to a formal and pre-defined evaluation criterion that addresses three different aspects:

1. Quality of Content;
2. Potential Effective as a Teaching Tool; and
3. Ease of use.

After peer-reviewers report their evaluations, the chief-editor composes one single report which is published in the repository with the authorization of the authors.

In addition to peer-reviewers evaluations, MERLOT also allows the community of users to provide comments and ratings for the materials, complementing its strategy of evaluation with an alternative and more informal mechanism. The ratings of both

---

3 http://www.commonsensemedia.org/
4 www.merlot.org
Moreover, MERLOT also allows users to bookmark the resources in the so-called Personal Collections, providing them a way of organizing their favorite materials according to their individual interests [20]. At last, MERLOT annually gives a special award (the MERLOT Classics Awards) for outstanding materials according to a program criterion of the disciplines (see Figure 7). All these evaluative metadata together are used to sort learning materials every time a user performs a search in the repository.

![Figure 7. The MERLOT repository (Arts discipline learning materials)](image)

MERLOT is particularly peculiar in the sense that ratings are gathered from two well defined and different groups of people (general public and experts), which possibly come from distinct backgrounds and may have divergent opinions with respect to quality. In fact, these differences between reviewer’s groups can be considered as a strong point of the adopted approach, which provides complementary views about the same subject. In the next subsection we briefly describe the main characteristics and differences between these two approaches.

### 2.6 Peer-Review and Public-Review

Peer-review is conventionally known as the process of assessing a scientific paper or project idea by critical examination of third parties that are experts in the same work domain. This system is widespread in the process of publishing papers in journals and conferences, where the work under evaluation is submitted to a chief-editor which requests a group of fellow-experts to review it in order to obtain advices about whether or not the article must be accepted for publishing, and what further work is still required in the case of acceptance [21]. In the most widely adopted form of peer-review, the identity of the reviewers is hidden from the authors, as well as from the other reviewers. The defenders of peer-reviewing claim that this kind of professional approval serves as a way of assuring the quality of papers published. However, the system is not free from criticisms and issues such as: conflicts of interest, biases of the peers, unnecessary time delay, and the inability on detecting frauds, all mentioned as possible shortcomings of the peer-review process [22]. In any case, and despite the controversies regarding its efficiency, the peer-review system remains as the cornerstone for quality assurance in the academic field, and has also entered in the scene of educational resources after its implementation in MERLOT.

On the other hand, public-review is widely diffused in areas such as online vendors (e.g. Amazon, eBay) and several communities of interest (e.g. IMDb, YouTube, RYM, etc). In these, users normally benefit themselves from comments and ratings given by the community through the use of recommender systems (such as collaborative filters) which, based on the comparison of user’s profiles and the correlation of personal tastes, provide personalized recommendation of items and products that will probably be of their interest [23]. In this kind of social systems, the motivations and goals behind the users’ participation vary significantly, from the desire and need of social interaction, to professional self expression and reputation benefits [24]. Table 1 explores some other aspects which normally differentiate standard peer-review and public-review systems.

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Peer-Review</th>
<th>Public-Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator background</td>
<td>Expert in the field domain</td>
<td>Non-expert</td>
</tr>
<tr>
<td>Existence of official criteria or metrics</td>
<td>Yes</td>
<td>No/Sometimes</td>
</tr>
<tr>
<td>Community of evaluators</td>
<td>Restricted</td>
<td>Wide opened</td>
</tr>
<tr>
<td>Common models</td>
<td>Pre-publication</td>
<td>Post-publication</td>
</tr>
<tr>
<td>Domain</td>
<td>Scientific field, journals and funding calls</td>
<td>Online vendors, communities of interest</td>
</tr>
<tr>
<td>Motivation</td>
<td>Prestige, fame, to determine the quality and direction of research in a particular domain, obligation</td>
<td>Desire and need of social interaction, professional self expression, reputation</td>
</tr>
<tr>
<td>Communication among evaluators</td>
<td>Not allowed</td>
<td>Encouraged</td>
</tr>
<tr>
<td>Selection of evaluators</td>
<td>Editor responsibility</td>
<td>None</td>
</tr>
<tr>
<td>Financial compensation</td>
<td>Normally none</td>
<td>None</td>
</tr>
<tr>
<td>Time taken for the evaluation</td>
<td>Typically slow</td>
<td>Typically fast</td>
</tr>
<tr>
<td>Level of formality</td>
<td>Formal process for editing and revision</td>
<td>Informal</td>
</tr>
<tr>
<td>Author’s identity</td>
<td>Masked</td>
<td>Non-masked</td>
</tr>
<tr>
<td>Requirements to be a reviewer</td>
<td>To be an expert in the field and to be invited</td>
<td>Creation of a member’s account</td>
</tr>
</tbody>
</table>
3. Existing strategies for evaluation inside LORs versus automated evaluation

Although current strategies for evaluation inside repositories can be considered successful at some extent, the amount of learning objects inside repositories tend to grow faster than the capacity of the community to evaluate them [25]. Such condition turns impractical to rely only on human effort to classify good quality resources as it becomes impossible to provide evaluative metadata for every single resource in the repository, thus leaving many resources of the current repositories without any measure of quality at all. This situation has raised the concern about the development of new automated techniques and tools that could be used to complement the existing approaches in order to relieve manual work. The actual abundance of resources inside repositories [26] and the availability of contextual evaluations in some of them have opened the possibility of seeking for intrinsic metrics of learning objects that could be used as indicators of quality. This means to say that learning objects could be “mined” and quantitative measures of good and not-good resources could be compared in order to discover intrinsic attributes associated with quality, thus allowing the creation of statistical profiles of good and poor resources that could serve as the basis for quality prediction.

Even though automated analysis cannot replace traditional inspection techniques, it carries the potential of offering an inexpensive and time saving mechanism to a priori explore the quality of materials, therefore complementing other existing approaches. The deployment of such automated tools would certainly improve the general quality of the services provided by repositories regarding the processes of searching, selecting and recommending good quality materials. Contributors could, for instance, benefit of such new feature by evaluating beforehand the quality of their resources, which would allow their improvement through the use of the quality metrics referenced by the tool.

Initial works in this direction have been developed by [27] who proposed a complementary approach for automated evaluation that relies on the data that can be directly extracted from the learning resources themselves in combination with evaluative metadata. The main advantage of such proposal is to offer a tool which is able to assess quality of new resources inserted in the repository without the need of annotations about them. The authors have offer the very first foundations for the development of such tool by contrasting intrinsic metrics of highly-rated and poorly-rated learning objects stored in MERLOT repository and identifying which metrics are mostly associated with rated resources in the context of that repository. In their work, they have found that the tested metrics present different profiles and tendencies between good and not-good materials depending on the category of discipline and the type of material to which the resource belong. For instance, positive correlations were found between the Number of Images and highly rated learning resources in the disciplines of Education, Mathematics and Statistics, and Science and Technology, and for the Number of Applets in the Business discipline [27]. Moreover they built a Linear Discriminant Analysis model based on the metrics which was able to distinguish between good and not-good materials (for the discipline of Science and Technology and Simulation type) with an overall accuracy of 91.49%, a remarkable achievement for a preliminary attempt towards automated evaluation.

In another experiment [28], the authors tested a slightly different and more algorithmic approach, i.e., the models were generated exclusively through the use of data mining algorithms. Among other good results, one can mention the model for Humanities \( \cap \) Simulation that was able to classify good resources with 75% of precision and not-good resources with 79%; and the model developed for Mathematics \( \cap \) Tutorial with 79% of precision for classifying good resources and 64% for classifying not-good ones.

The same approach was tested for the Connexions repository in [29], however the generated models presented poor performances for classifying resources. According to the authors, this may be a consequence of the small size of the population of resources that had evaluative metadata (endorsements). Therefore it is still needed to wait the growth of endorsements in the repository in order to better evaluate the feasibility of creating models for automatically classify resources according to their amount of endorsements.

Whether or not the methodology can be extrapolated for other repositories is still a subject for further investigation and research. In the mentioned works, the authors relied on information (categories of discipline, types of materials, peer reviewers and users ratings, endorsements) that are not (necessarily) available in other learning resources repositories. In the cases where some of these information are not available, alternative ways of searching for LOs quality must be found in order to contrast with the metrics for the establishment of these profiles, such as, for instance, the use of ranking metrics [30] or other kinds of evaluative metadata available in such repositories.

4. Final Remarks

Evaluating quality of learning objects is a difficult task that normally involves several distinct aspects and different stakeholders, and the existing learning object evaluation methods and frameworks are not free from ambiguities. Different Learning object repositories are frequently adopting strategies that rely on the community of users and experts that assess the quality of the resources by rating and commenting them. Such evaluations can be performed according to a formal and pre-defined evaluation criterion that addresses specific different aspects, or in a more informal way and without pre-defined specifications. The resulting set of evaluations is then used by LORs to facilitate the process of searching and ranking resources and is considered as a social body of knowledge that serves as an external memory aid for individuals that navigate in such portals. The existence of such evaluations also opens de possibility for the future implementation of personalized recommendations based on the preferences of the users [31]. At the same time that current strategies have established themselves as the main alternative for quality evaluation inside repositories, they are still insufficient to cover the huge amount of resources that continuous grown in such platforms. Therefore, there is an urgent need for the development of alternatives that help to boost the provision of quality information in complement of existing manual strategies.
5. ACKNOWLEDGMENTS
This work has been funded by CYTED (Ibero-American Programme for Science, Technology and Development) as part of project “RIURE - Ibero-American Network for the Usability of Learning Repositories” (code 513RT0471), and by the Núcleo de Estudios e Investigaciones en Educación Superior del Sector Educativo del MERCOSUR (Center of Studies and Investigations on Higher Education of MERCOSUL) as part of Project “Marcos regulatorios, Modelos institucionales y formatos didácticos: Dimensiones de la calidad en las Prácticas de Enseñanza a Distancia en países del MERCOSUR” (Regulatory Marks, Institutional Models and Didactic Forms: Quality Dimensions of Distance Learning Practice in MERCOSUL Countries).

6. REFERENCES


