

Towards a Learning Object pedagogical quality metric based on the LORI evaluation model

Aldo Gordillo, Enrique Barra and Juan Quemada

Escuela Técnica Superior de Ingenieros de Telecomunicación

Universidad Politécnica de Madrid

Avenida Complutense 30, 28040, Madrid, Spain

{agordillo, ebarra, jquemada}@dit.upm.es

Abstract— Evaluating and measuring the pedagogical quality of Learning Objects is essential for achieving a successful web-based education. On one hand, teachers need some assurance of quality of the teaching resources before making them part of the curriculum. On the other hand, Learning Object Repositories need to include quality information into the ranking metrics used by the search engines in order to save users time when searching. For these reasons, several models such as LORI (Learning Object Review Instrument) have been proposed to evaluate Learning Object quality from a pedagogical perspective. However, no much effort has been put in defining and evaluating quality metrics based on those models. This paper proposes and evaluates a set of pedagogical quality metrics based on LORI. The work exposed in this paper shows that these metrics can be effectively and reliably used to provide quality-based sorting of search results. Besides, it strongly evidences that the evaluation of Learning Objects from a pedagogical perspective can notably enhance Learning Object search if suitable evaluations models and quality metrics are used. An evaluation of the LORI model is also described. Finally, all the presented metrics are compared and a discussion on their weaknesses and strengths is provided.

Keywords— learning object; metric; pedagogical quality; LORI

I. INTRODUCTION

Learning Objects (LOs) are reusable digital resources tagged with metadata that are self-contained and that can be used for education. These resources are increasingly used both in traditional classrooms as well as computer-based learning environments due to their great benefits. LOs facilitate reuse and interoperability [1], leading to minimization of production cost, time saving and the quality enhancement of digital learning experiences [2]. The multiple benefits of using LOs have been exposed by several empirical researches ([3], [4], [5], [6], [7]) that have examined their instructional effectiveness and learning outcomes across different educational environments.

Today, thousands of LOs can be freely searched with standardized metadata through online educational repositories, also called Learning Object Repositories (LORs), facilitating educators and students to obtain new materials. However, due to the emergence of large repositories and the prevalence of low quality materials there are still some barriers that need to be overcome to extend the effective use of these objects. First off, teachers need some assurance of value and pedagogical quality of the LOs before making them part of the curriculum.

Secondly, LORs need some mechanism to measure the quality of the LOs in order to guarantee a minimum quality level of the published educational resources. Lastly, there is a need to include a pedagogical quality dimension into the ranking metrics used by search engines in order to save users time when searching for high quality resources. Some studies [8], have pointed out that teachers can take over an hour to find an appropriate LO, and that several of them can note the search process to be time consuming and occasionally frustrating. With the aim of addressing this problem, some LORs provide a metadata-based search. Nevertheless, this approach usually does not take into account pedagogical quality information. Besides, empirical studies have shown that common users found this search approach too difficult [9]. Therefore, most repositories provide a “Simple Search” approach, where users express their information needs with keywords or query terms [10]. All these barriers point out a clear need for evaluation models and metrics to measure the pedagogical quality of LOs.

As a consequence of this need, several evaluation models have been proposed to evaluate the pedagogical quality of LOs. The most cited and one of the most used and tested model for evaluating LO quality from a pedagogical perspective is LORI [11], which stands for Learning Object Review Instrument. Based on these models, it is possible to define metrics to measure the pedagogical quality of the LOs. Pedagogical quality metrics can be used to grade the quality of LOs (e.g. [12]), to build ranking metrics in order to provide quality-based sorting of search results (e.g. [13]), or to generate enhanced recommendations (e.g. [14]). Nevertheless, the quality metrics used in practice usually consist simply of a sum of the partial scores corresponding to each criterion considered by the model, assuming that all of them are equally important. Given that, it is reasonable to think that better metrics could be defined by assigning different importance to each criterion and/or by establishing different relations among them. Besides, no much effort has been put in exploring the characteristics of different LO pedagogical quality metrics in order to compare them and discover the best one for each scenario.

This paper proposes and evaluates a set of LO pedagogical quality metrics based on the LORI evaluation model, showing that these metrics can be used to provide quality-based sorting of search results in a reliable and effective way. Besides, all the presented metrics are compared, and a detailed discussion is provided on their weaknesses and strengths.

The rest of the paper is organized as follows. The next section reviews related work of LO evaluation and metrics. Section 3 provides an overview of LORI. Section 4 exposes an evaluation of the LORI model, in which the evaluations and feedback from a team of 15 reviewers were used to determine two different sets of weights for the LORI criteria. Section 5 presents the proposed quality metrics. The evaluation and comparison of the different quality metrics is covered by section 6. Section 7 provides some discussion. Finally, section 8 finishes with some concluding remarks and future work.

II. RELATED WORK

There are many evaluation models that aim to evaluate the quality of LOs. These models can be classified depending on their concept of LO quality. Some of them evaluate metadata quality (e.g. [15], [16]) or reusability (e.g [17], [18]). Other models evaluate the whole LO including a content quality evaluation from a pedagogical perspective. Examples of these models are among others LORI [11], LOEM [19], the MERLOT model [20], WBLT-S and WBLT-T [21], LOEI [22] and MECOA [23]. There are also hybrid models that combine different quality indicators (e.g. [13]). Pedagogical evaluation models may also be characterized in terms of the criteria they define to evaluate LO quality and the type of the evaluations. Some models provide qualitative evaluations (e.g. reviews), while others provide quantitative evaluations (e.g. ratings), quality certificates, or a combination of several of them (e.g. ratings + comments). Among the ones that provide quantitative evaluations, some only provide ratings for each criterion or subset of criteria, while others define an overall quality metric in order to calculate an overall score enabling quick comparison among evaluated resources and quality-based sorting of search results.

Several metrics for LOs have been defined to measure not only pedagogical quality, but also to measure other aspects like metadata quality (e.g. [15], [24]), reusability (e.g. [17],[18]), popularity (e.g. [25]) or similarity (e.g. [25]). Other important metrics for LOs are the ranking metrics (e.g. [10], [13]), which usually combine several indicators of quality (e.g. pedagogical and metadata quality, popularity) and relevance (personal and situational context) to sort search results in LORs. The metrics that measure aspects of LO quality are termed LO quality metrics. Some of them may act as individual indicators of some aspect of LO quality (e.g. reusability) while others may provide an overall quality measure. Moreover, the same as the ranking metrics, they can be built by combining several other metrics. A quality metric is characterized by the evaluation model and/or criteria it takes into account, the mathematical process it uses to calculate the measure of LO quality, and the scale in which the final numeric value is yielded.

A few metrics have been defined upon LO evaluation models to measure the pedagogical quality of LOs. For instance, in the MERLOT repository [20], resources are graded by users and appointed peer reviewers with comments and ratings on a 5-point scale according to three criteria: content quality, effectiveness as a learning tool and usability. The overall quality metric defined as the equally weighted mean of the three criteria is used to provide quality-based sorting of the search results. The University of Information Sciences in Cuba

defined a LO evaluation model together with some quality metrics in order to measure the level of quality of their learning resources [12]. The model considers three aspects: pedagogical (15 criteria), design (9 criteria) and technological (11 criteria). Each of these 35 criteria is rated in a 0-4 scale. The model also defines a quality metric for each of the aspects and another one based on them to calculate the overall quality score. Depending on the scores obtained according to these metrics, LOs may be considered “Unsuitable”, “Not very suitable”, “Suitable” or “Very Suitable”. The quality metrics consider all criteria equally important, but LOs must meet a minimum threshold score on each of the three individual aspects to be accepted or to climb on the assessment scale. Other example of quality metric is the one used in eLera [26], a website that provides communities of teachers, learners and developers with a set of web-based tools for collaborative evaluation of LOs based on the LORI evaluation model. In eLera, the overall scores of the LOs used to sort the search results are calculated by averaging the ratings over LORI criteria and reviewers.

III. LORI (LEARNING OBJECT REVIEW INSTRUMENT)

Reviewers can use LORI to assess LOs from a pedagogical perspective according to certain criteria. The last version is LORI 1.5 [11], which considers the nine criteria (or LORI items) listed in Table I.

TABLE I. ITEMS IN LORI 1.5

LORI Item	Brief Description
1. Content Quality	Veracity, accuracy, balanced presentation of ideas, and appropriate level of detail
2. Learning Goal Alignment	Alignment among learning goals, activities, assessments, and learner characteristics
3. Feedback and Adaptation	Adaptive content or feedback driven by differential learner input or learner modeling
4. Motivation	Ability to motivate and interest an identified population of learners
5. Presentation Design	Design of visual and auditory information for enhanced learning and efficient mental processing
6. Interaction Usability	Ease of navigation, predictability of the user interface, and the quality of the interface help features
7. Accessibility	Design of controls and presentation formats to accommodate disabled and mobile learners
8. Reusability	Ability to use in varying learning contexts and with learners from different backgrounds
9. Standards Compliance	Adherence to international standards and specifications

For each criterion, reviewers can enter comments and ratings on a 5-point scale ranging from 1 (low) to 5 (high). Reviewers can skip criteria that they are unable to assess or that they do not feel qualified enough to rate by selecting the “Not applicable” option instead of a numeric rating. This can also be done for criteria that are judged not relevant to the LO. So, LORI provides qualitative evaluations (comments) and quantitative evaluations (ratings). In addition to specifying a rating for each criterion, LORI 1.5 proposes an overall quality metric that consists of the average rating of all criteria. The LORI evaluation model is intended to cover a wide range of educational settings, but it is worth pointing out that only evaluates LOs from the reviewers’ perspective.

LORI may be used for either individual or group reviews. When it is used for collaborative evaluation, it is recommendable to use the convergent participation model [27]. In this model, small teams of reviewers are formed from participants representing relevant knowledge sets (e.g. subject matter experts or e-Learning professionals). First, each member of the team evaluates the LOs individually. Then, all members meet, usually via web videoconference, to compare and discuss their evaluations and establish a final team evaluation.

LORI has been tested in a few studies ([28], [29]), which have shown that it can be used to reliably assess some aspects of LOs. Nevertheless, other studies have found some limitations of the model. For instance, [30] pointed out that LORI evaluations could be not sufficient to predict learning outcomes of students in some scenarios.

There are web tools developed to facilitate the evaluation of LOs using LORI. The most known are the collaborative evaluation tools provided by eLera [26], the website for which LORI was developed. Other example is webLORI [31], which is basically an online version of LORI. There is also an open source platform called LOEP [32], which aims to facilitate LO evaluation in different scenarios by supporting several evaluation models (including LORI) and quality metrics.

IV. LORI EVALUATION

As a first step to define new pedagogical quality metrics based on LORI, we conduct a study to evaluate some aspects of the evaluation model. The main goal of this evaluation was to determine the importance or weight that should be given to each of the LORI items. Although some studies (e.g. [31]) have proposed further research on this question, this issue has not been addressed yet.

The participants of this study were the 15 members of a team of reviewers. This team used an online version of LORI 1.5 to evaluate the LOs of a Learning Object Repository called ViSH (<http://vishub.org>) [33]. All these LOs were created by teachers, researchers and students using an online e-Learning authoring tool called ViSH Editor [34], available on the ViSH site. This authoring tool allows to create interactive presentations using different types of resources such as text, images, videos, websites, SCORM packages, flash objects, quizzes and PDF files among others. Therefore, the evaluated LOs were notably different and quite varied in quality. The web-based tool used to evaluate the LOs was provided by the LOEP platform mentioned above. Thus, reviewers evaluated the LOs using web forms. An optional field was added to the web forms in order to allow reviewers to propose an overall score for the LOs together with their evaluations.

At the beginning of the study, reviewers studied the LORI user manual and received a 4-hour training session on how to evaluate LOs using LORI and how to use the LOEP platform to carry out the evaluations. Email support was available throughout the duration of the study. After that, a set of 209 LOs of the ViSH repository was equally shared out between reviewers. This sharing out was done automatically using the LOEP platform, taking into account several factors like the topic of the LOs and the areas of expertise of the reviewers in order to assign to each reviewer the most suitable LOs.

Over four months, the team of reviewers evaluated the 209 LOs with LORI 1.5, generating a total amount of 740 evaluations. LORI was used for individual reviews, but each LO was evaluated by at least 3 different reviewers (3.59 evaluations per LO on average). After these four months, a survey among the team members was conducted to collect feedback about LORI.

The sample of this study consisted of 15 reviewers, 9 males (60%) and 6 females (40%), 23 to 43 years of age ($M=30.8$, $SD=5.6$). The team was formed by 9 educators, 4 e-Learning professionals and 2 designers. Table II summarizes some results of the survey.

TABLE II. LORI SURVEY RESULTS

Please indicate your level of agreement on each of the following statements about LORI [1 (strongly disagree) - 5 (strongly agree)]	Reviewers (n=15)	
	M	SD
LORI is an excellent tool for evaluating the quality of Learning Objects	4.2	0.8
The documentation (i.e. the LORI user manual) is truthful, accurate, well presented, and has the appropriate level of detail	4.1	0.9
The training session on LORI was truthful, accurate, well presented, and had the appropriate level of detail	3.9	0.7
I have improved my ability to judge the quality of LOs after using LOEP	4.0	0.9
Please indicate your comfort level rating Learning Objects on the following LORI items [1 (Very uncomfortable) - 5 (Very comfortable)]	M	SD
1. Content Quality	4.2	0.7
2. Learning goal alignment	4.1	1.1
3. Feedback and adaptation	3.9	1.0
4. Motivation	4.3	0.8
5. Presentation design	4.4	0.6
6. Interaction usability	3.9	0.6
7. Accessibility	2.9	1.0
8. Reusability	3.1	1.2
9. Standards compliance	3.3	1.3

Reviewers perceived LORI as a useful tool for evaluating LO quality. Furthermore, they indicated that the provided documentation and the training session were helpful to perform the evaluations. Lastly, they felt that they have improved their ability to judge LO quality after this experience.

Regarding the comfort level rating the different LORI items, reviewers felt more comfortable rating the items related to pedagogical criteria (items 1 to 6) than rating the items more related to technological criteria (items 7 to 9). The explanation for this is that non-technical reviewers have difficulties to judge items related to technological criteria. This fact becomes clear when we divide the results in two groups: technicians (e-Learning professionals) and non-technicians (educators and designers). Table III exposes these results, showing that reviewers without technical knowledge do not felt very comfortable rating LORI items related to technological aspects such as Accessibility ($M=2.5$, $SD=0.8$), Reusability ($M=2.8$, $SD=1.1$) and Standards Compliance ($M=2.9$, $SD=1.3$).

TABLE III. LORI SURVEY RESULTS (II)

Please indicate your comfort level rating Learning Objects on the following LORI items	Technicians (n=4)		Non-Technicians (n=11)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
7. Accessibility	4.0	0.8	2.5	0.8
8. Reusability	3.8	1.5	2.8	1.1
9. Standards Compliance	4.5	0.6	2.9	1.3

The fact that 27.1% of the evaluations carried out by educators lacked of some of these three items, reinforces the idea that usually non-technical reviewers are unable or do not feel qualified enough to judge these criteria. This conclusion is consistent with the findings of other studies. For instance, in [28], most of the participants recognized a lack of knowledge of the metadata and accessibility standards. Similarly, in [29], the original LORI items that dealt with accessibility and standards compliance were removed because the participants lacked the necessary knowledge to provide an assessment in both of those areas.

According to the LORI guidelines, it could be argued that the Reusability item is not a pure technological criterion since it is focused on educational portability. However, in practice, the ability of a LO to be reused in varying learning contexts and with learners from different backgrounds strongly relies on technological aspects. Reviewers need a deep knowledge about the different technologies in which the resources are provided in order to suitably judge to what extent a LO can be customized (e.g. using authoring tools), translated or integrated into other resources or e-Learning systems, and how much effort is needed to do so. The empirical data shown in this study sustain this fact.

In conclusion, to reliably evaluate a LO with LORI, it should not only be reviewed by educators with knowledge of its subject, but also by at least one e-Learning professional or technician. In our study, each LO was evaluated by at least one e-Learning professional and one educator with knowledge of the LO subject. These results also suggest that future versions of LORI should consider replacing the Reusability item for two different items: "Pedagogical Reusability" and "Technological Reusability". This way, educators could focus on educational portability while technicians will take care of the technological constraints. The overall reusability score could be calculated based on these two items in order to offer a reliable and truthful measure of reusability.

A. Collected Weights

Reviewers were also asked about which LORI items should be weighed more. For each one of the LORI items, reviewers rated the importance that it has for them regarding pedagogical quality on a 0 to 10 scale, being 0 valueless and 10 extremely valuable. The weights were averaged over all reviewers and were normalized to sum to one. The list of obtained weights, which hereafter we will call "Collected Weights", is shown in Table IV.

TABLE IV. LORI COLLECTED WEIGHTS

LORI Item	Weight
1. Content Quality	$W_1 = 0.1724$
2. Learning Goal Alignment	$W_2 = 0.1207$
3. Feedback and Adaptation	$W_3 = 0.1138$
4. Motivation	$W_4 = 0.1414$
5. Presentation Design	$W_5 = 0.1379$
6. Interaction Usability	$W_6 = 0.1034$
7. Accessibility	$W_7 = 0.0655$
8. Reusability	$W_8 = 0.0759$
9. Standards Compliance	$W_9 = 0.0690$

B. Inferred Weights

As was explained before, reviewers had the option to propose an overall score for the LOs together with their evaluations. A total of 241 evaluations (32.6%) out of 740 included a proposed overall score. Based on these scores and the ratings of the items, we used a least-squares multiple linear regression analysis to calculate a set of weights for the LORI items. In this case, the proposed score was considered the dependent (or response) variable and all the ratings of the items were the independent (or explanatory) variables. The coefficients (i.e. the weights) that produced the best fit of the weighted arithmetic mean of all item ratings against the proposed score were estimated.

All LOs of this study were created using the ViSH Editor authoring tool and published in the ViSH repository. That implies that all of them were SCORM compliant LOs, with W3C-compliant HTML5 code, and that their metadata were available in the IEEE LOM format. All LOs were tagged with, at least, essential metadata since these data are mandatory for publishing LOs in ViSH. For these reasons, all LOs in this study received ratings between 4 and 5 in the Standards Compliance item according to the LORI guidelines. The rest of the items received ratings between 1 and 5, covering the full range of possible scores. So, given that there were no significant differences regarding standards compliance among the LOs, and that the range covered was too narrow to obtain a reliable weight estimation for the Standards Compliance item using a multiple linear regression analysis, we decided to set this weight to the value obtained in the Collected Weights: $W_9=0.069$. Thereby, we used the least-squares multiple linear regression analysis to calculate the weights for the LORI items 1 to 8, approximating W_9 to a value of 0.069. Table V shows the obtained "Inferred Weights".

V. PEDAGOGICAL QUALITY METRICS

This section describes several pedagogical quality metrics for LOs based on the LORI evaluation model. All the metrics exposed here have been implemented in the LOEP platform. Thereby, it is possible to automatically calculate scores according to these quality metrics for LOs evaluated with LORI 1.5.

TABLE V. LORI INFERRED WEIGHTS

LORI Item	Coefficient	p-value	Weight
1. Content Quality	0.3828	< 0.0001	$W_1 = 0.1475$
2. Learning Goal Alignment	0.1724	0.006	$W_2 = 0.0665$
3. Feedback and Adaptation	0.3588	< 0.0001	$W_3 = 0.1383$
4. Motivation	0.6299	< 0.0001	$W_4 = 0.2427$
5. Presentation Design	0.0141	0.8	$W_5 = 0.0054$
6. Interaction Usability	0.3436	< 0.0001	$W_6 = 0.1324$
7. Accessibility	0.2334	< 0.001	$W_7 = 0.0899$
8. Reusability	0.2814	< 0.0001	$W_8 = 0.1084$
9. Standards Compliance	-	-	$W_9 = 0.0689$
Regression Statistics			
Observations	Multiple R	R Square	Significance F
241	0.9934	0.9868	< 0.0001

A. LORI Arithmetic Mean (LORI AM)

We start describing the simplest quality metric, which assumes that all items are equally important. This is the metric used by eLera [26] to sort the search results. The equation that calculates the LO score takes an input score vector $\{i_1, \dots, i_9\}$, being i_i the score of the LORI item i in a scale from 1 to 5, and yields a single real value in a scale from 0 to 10 according to the following expression:

$$s(\{i_1, \dots, i_9\}) = \frac{5}{18} \times \sum_{i=1}^9 (i_i - 1), \quad s(\{i\}) \in [0,10] \quad (1)$$

All the presented metrics yield a single real value in a 0 to 10 scale in order to allow comparison between scores. The i_i variable corresponds to the rating of the LORI item i given by the reviewers. In order to calculate the overall LO score, the i_i variable should be the average of the ratings of the LORI item i given by all reviewers.

B. LORI Weighted Arithmetic Mean (LORI WAM)

This metric calculates the score of a LO as the weighted arithmetic mean of all LORI items scores, giving different importance to each criterion. This general metric is characterized by a weights vector $\{W_1, \dots, W_9\}$, where W_i corresponds to the weight assigned to the LORI item i . Therefore, different weights vectors lead to distinct quality metrics. In fact, the LORI AM metric is a particular case of this metric when $W_i = 1/9 \forall i$ belonging to $\{1 \dots 9\}$. The equation that calculates the LO score takes an input score vector $\{i_1, \dots, i_9\}$, being i_i the score of the LORI item i in a scale from 1 to 5, and yields a single real value in a scale from 0 to 10 according to the following expression:

$$s(\{i_1, \dots, i_9\}) = \frac{5}{2} \times \sum_{i=1}^9 W_i \times (i_i - 1), \quad s(\{i\}) \in [0,10] \quad (2)$$

where $\sum_{i=1}^9 W_i = 1$ and $W_i \geq 0 \forall i \in \{1, \dots, 9\}$

Based on this general expression, we define two quality metrics using the weights obtained in the previous section. We will refer to the LORI WAM metric that uses the Collected Weights (Table IV) as *LORI WAM CW*, and we will refer to the LORI WAM metric that uses the Inferred Weights (Table V) as *LORI WAM IW*.

C. Dividing Pedagogical and Technological Criteria

LORI items can be divided in two subsets: a subset of pedagogical items that covers items 1 to 6, and a subset of items related to technological criteria that includes the items 7, 8 and 9. This fact has been previously described in section 4, showing also that technological items should be rated by reviewers with deep technical knowledge such as e-Learning professionals. For each one of these two subsets, we defined a new quality metric by using the LORI WAM metric and specifying a new weights vector. The Collected Weights were used as a base for defining the new weights vectors, which are shown in Table VI.

1) LORI Pedagogical WAM (LORI PWAM)

In this case the weights W_1 to W_6 were normalized to sum to one (see Table VI). The equation that calculates the LO score takes an input score vector $\{i_1, \dots, i_6\}$, being i_i the score of the LORI item i in a scale from 1 to 5, and yields a single real value in a scale from 0 to 10.

$$s(\{i_1, \dots, i_6\}) = \frac{5}{2} \times \sum_{i=1}^6 W_i \times (i_i - 1), \quad s(\{i\}) \in [0,10] \quad (3)$$

where $\sum_{i=1}^6 W_i = 1$ and $W_i \geq 0 \forall i \in \{1, \dots, 6\}$

2) LORI Technological WAM (LORI TWAM)

In this case the weights W_7 , W_8 and W_9 were normalized to sum to one (see Table VI). The equation that calculates the LO score takes an input score vector $\{i_7, i_8, i_9\}$, being i_i the score of the LORI item i in a scale from 1 to 5, and yields a single real value in a scale from 0 to 10.

$$s(\{i_7, i_8, i_9\}) = \frac{5}{2} \times \sum_{i=7}^9 W_i \times (i_i - 1), \quad s(\{i\}) \in [0,10] \quad (4)$$

where $\sum_{i=7}^9 W_i = 1$ and $W_i \geq 0 \forall i \in \{7, 8, 9\}$

TABLE VI. WEIGHTS FOR LORI PWAM AND TWAM METRICS

LORI Item	LORI PWAM Weights	LORI TWAM Weights
1. Content Quality	$W_1 = 0.2183$	-
2. Learning Goal Alignment	$W_2 = 0.1529$	-
3. Feedback and Adaptation	$W_3 = 0.1441$	-
4. Motivation	$W_4 = 0.1791$	-
5. Presentation Design	$W_5 = 0.1746$	-
6. Interaction Usability	$W_6 = 0.1310$	-
7. Accessibility	-	$W_7 = 0.3113$
8. Reusability	-	$W_8 = 0.3607$
9. Standards Compliance	-	$W_9 = 0.3280$

3) Defining Overall Quality Metrics

Based on the LORI PWAM and LORI TWAM metrics, we can build new quality metrics by establishing non-linear relationships between pedagogical and technological items. These metrics are particular cases of the following general expression:

$$s(\{i_1, \dots, i_9\}) = f(S_{PWAM}(\{i\}), S_{TWAM}(\{i\})), \quad s(\{i\}) \in [0,10] \quad (5)$$

D. LORI Orthogonal Metric

This metric represents LORI PWAM and LORI TWAM as two orthogonal vectors in a two-dimensional coordinate system, and calculates the LO score as the modulus of the vector sum of these vectors in a 0-10 scale.

$$s(\{i_1, \dots, i_9\}) = \frac{\sqrt{2}}{2} \times \sqrt{(S_{PWAM})^2 + (S_{TWAM})^2}, \quad s(\{i\}) \in [0,10] \quad (6)$$

E. LORI Logarithmic Metric

This metric defines a logarithmic relationship between LORI PWAM and LORI TWAM according to the following expression:

$$s(\{i_1, \dots, i_9\}) = \frac{S_{PWAM} \times \ln(A \times S_{TWAM} + 1)}{\ln(10 \times A + 1)} \quad (7)$$

$A > 0, \quad s(\{i\}) \in [0,10]$

The key idea of this metric relies on the fact that technology features of a LO (e.g. accessibility or standards compliance) can gradually increase its overall quality, but only to the extent that it has a high pedagogical quality. This way, an extremely poor resource (e.g. a SCORM package with a blank page) will be rated with 0 regardless of its metadata or supported standards. This metric also penalizes resources with extremely low scores in the technological criteria. This way, in the extreme case that a resource cannot be accessed by any common device and is not compliant with any standard, it will be also rated with a 0 regardless its content. The first points in the technological scale are given more importance than the last ones. That makes sense since for instance, essential metadata (e.g. title, description, language, keywords) are much more important than other metadata fields. When a LO achieve a minimum technical specifications, the added value of incorporating new features is decreased. The parameter A enables to adjust the curve and set the desired quality thresholds. In this study, we set A to a value of 2.

F. LORI Square Root Metric

This metric defines a square root relationship between LORI PWAM and LORI TWAM according to the following expression:

$$s(\{i_1, \dots, i_9\}) = \sqrt{S_{PWAM} \times S_{TWAM}}, \quad s(\{i\}) \in [0,10] \quad (8)$$

This metric is quite similar to the Logarithmic Metric, since it penalizes LOs with extremely low technological or pedagogical scores. In this case, there is no parameter to adjust the curve, but for the values $S_{PWAM}=5$ and $S_{TWAM}=5$, the yielded overall score is 5.

VI. METRICS EVALUATION AND COMPARISON

All the presented metrics are based on the same evaluation model (LORI). That implies that all of them rely on the same criteria (LORI items) and raw data (the LORI evaluations) to calculate the quality scores. Besides, all of them yield the final value in a 0 to 10 scale. However, the mathematical expressions used to calculate the final score according to the ratings of the LORI items are different in each metric. This may entail significant consequences in various aspects such as the ability of the metric to sort LOs for users according to their perceptions of quality, the way in which the scores are distributed among groups of LOs, or the reliability of the metric in order to effectively distinguish high quality from low quality LOs. To address this question, we conducted another study to evaluate and compare the different quality metrics.

This study consisted of an experiment in which participants were asked to simulate a search for physics teaching resources in a LOR. A total of 10 teachers participated in the study. A list of 12 randomly sorted LOs was presented to the participants as the results of this fictitious search. The 12 LOs were selected among the 209 LOs of ViSH that had been evaluated with LORI 1.5 by the reviewers team. All LOs were about physics. LOs with diverse quality were selected with the aim of cover the maximum possible range of scores. On the other hand, some LOs with very similar quality were also selected to reliably measure the accuracy of the metrics. Firstly, the participants were required to rate each of the 12 LOs of the list in terms of their pedagogical quality in a scale from 0 (lowest) to 10 (highest). Participants were also requested to give ratings lower than 5 to the LOs they feel that are not of high enough quality to be used as educational resources. After that, the participants were required to reorder the list of LOs according to their pedagogical quality, placing the LO with higher quality in the first position. This way, a human-generated ranking of the LOs was obtained. Then, we used all the presented metrics to reorder the list. Finally, the reordered lists were compared against the human-generated ranking. In order to measure the difference between the human-generated ranking and each of the rankings obtained using the metrics, we used the Kendall's tau (τ) coefficient [35], [36]. This metric measures the degree of correspondence between two rankings. If two rankings are identical, the τ coefficient is equal to 1, and if they are in inverse order, the τ coefficient is equal to -1. The τ coefficients obtained for each of the metrics are shown in Table VII.

In addition, other parameters are included in the Table VII. As a measurement of the dispersion of the LO scores, their standard deviation was also calculated. Besides, the covered range is shown, indicating the minimum and maximum scores obtained for each metric. The larger the standard deviation is and the wider the range is, the better is the metric distributing the scores. The score error was defined as the difference between the score calculated by the metric and the average score of the users. The smaller the error is, the more reliable is the metric. In the experiment, participants give ratings higher or equal to 5 to LOs they consider acceptable to be used for education. Similarly, metrics can define a quality threshold and accept LOs scored above it. In this study, we set all the quality thresholds to 5. The "Success rate in Acceptance" parameter shows the percentage of LOs for which the metric produces the

TABLE VII. METRICS COMPARISON

Metric	Equation	Rank Correlation		SD	Scores						Success rate in Acceptance	Accuracy
		Kendall's tau (τ)	2-sided p-value		Range			Error				
					Min	Max	%	Mean	SD	Max		
LORI AM	(1)	0.91	< 0.0001	2.44	1.18	9.17	79.9%	0.63	0.64	1.75	100%	1.55
LORI WAM CW	(2)	0.88	< 0.0001	2.65	0.75	9.34	85.9%	0.62	0.64	1.86	100%	1.55
LORI WAM IW	(2)	0.94	< 0.0001	2.63	0.81	9.40	85.9%	0.67	0.50	1.59	100%	1.55
LORI PWAM	(3)	0.88	< 0.0001	3.01	0.08	9.60	95.2%	0.72	0.68	2.02	100%	1.55
LORI TWAM	(4)	0.55	0.0164	1.56	3.25	8.97	57.2%	1.38	1.11	3.64	83.3%	5.56
LORI Orthogonal	(6)	0.91	< 0.0001	2.02	2.30	9.00	67.0%	0.90	0.82	2.12	100%	0.89
LORI Logarithmic (A=2)	(7)	0.85	< 0.001	2.85	0.05	9.06	90.1%	0.59	0.44	1.33	91.7%	1.55
LORI Square Root	(8)	0.76	< 0.001	2.48	0.51	8.96	84.5%	0.63	0.62	1.89	100%	1.44

same result as the users when deciding if a LO should be accepted. Finally, the “accuracy” parameter indicates the minimum difference that should exist between the average scores (of the users) of two LOs to ensure that the metric is capable of sorting the LO list with zero error. The smaller the accuracy is, the more capable will be the metric in sorting LOs with similar scores.

VII. DISCUSSION

Considering the obtained Kendall’s tau coefficients, we can state that all the presented metrics, except the LORI TWAM metric, do a great job at providing quality-based sorting of search results to educators. A ranking metric that not takes into account quality information of the LOs would record on average a Kendall’s tau coefficient of 0. Thus, these metrics clearly outperform this baseline metric, showing that the use of quality metrics based on LO evaluation can notably enhance LO search. There will always be differences between the automatic and human-generated rankings, since the metrics calculate the scores based on evaluations carried out by external reviewers, which judge the LOs from a different perspective than most end users. The efficacy of this approach to sort search results is directly dependent on the validity of the evaluation model used to generate the quality ratings [11]. Given that, we can state that LORI is a reliable model for evaluating LO quality.

The LORI TWAM metric, which is based on technological criteria, was found ineffective for sorting results. The larger value recorded by the accuracy parameter (5.56) denotes that this metric is unable to differentiate LOs according to their pedagogical quality. This fact points out that quality indicators such as metadata quality should not be used as measure of overall quality. Despite this metric is not very useful alone, it can be used to build more powerful quality metrics.

The LORI PWAM metric recorded good results in this study. That is comprehensible since most educators do not pay attention to technological criteria (or are not aware or their importance) when rating LOs. However, in other scenarios this metric could obtain much worse results if it is used alone.

These results show that using an average weighted mean with an appropriate set of weights for the criteria can bring better results than the typical LORI AM metric. The LORI WAM IW metric outperformed LORI AM on several parameters such as Kendall’s tau coefficient, scores range and maximum error. Anyway, the LORI AM metric achieves good results. The LORI Logarithmic and LORI Square Root metrics have been found useful to penalize LOs with extremely low technological or pedagogical scores. The Logarithmic metric records the smallest average and maximum error with a wide score range. The recorded success rate in acceptance suggests that the threshold for this metric should be slightly decreased. Although this study offers a preliminary analysis and evaluation of these metrics, a larger sample with different subsets of LOs should be used in further studies to analyze deeper the differences among these metrics.

VIII. CONCLUSIONS AND FUTURE WORK

This paper proposes a set of LO pedagogical quality metrics based on the LORI evaluation model. The work exposed in this paper shows that these metrics can be effectively and reliably used to provide quality-based sorting of search results. Besides, this work strongly evidences that the evaluation of LOs from a pedagogical perspective can notably enhance LO search if suitable evaluations models like LORI and quality metrics are used. Finally, a preliminary analysis of the differences among the presented LO quality metrics was offered, comparing several aspects such as ranking correlation coefficients, covered ranges of scores and accuracy.

The results of the evaluation of LORI suggest that to reliably assess a LO using LORI, it should be evaluated by at least one subject expert, and by a reviewer with deep technical knowledge. These results also suggest that future versions of LORI should seriously consider conducting an in-depth review of the reusability item.

Future work of this research will define new LO quality metrics, both for LORI (e.g. by giving different importance to reviewers’ ratings according to their suitability to evaluate the LO) and for other LO evaluation models. Finally, it would be extremely beneficial to compare LO quality metrics based on different evaluation models.

ACKNOWLEDGMENT

We wish to acknowledge our gratitude and appreciation to all the GLOBAL excursion project partners, and each one of the project team members, for their contribution to this paper.

REFERENCES

- [1] D. A. Wiley, "Learning Object Design and Sequencing Theory," Brigham Young University, 2000.
- [2] E. Duval, K. U. Leuven, and W. Hodgins, "A LOM Research Agenda," in *Proceedings of the 12th International World Wide Web Conference (2003)*, 2003, pp. 1–9.
- [3] R. McCormick and N. Li, "An evaluation of European learning objects in use," *Learning, Media and Technology*, vol. 31, no. 3, pp. 213–231, 2006.
- [4] S. Nurmi and T. Jaakkola, "Effectiveness of learning objects in various instructional settings," *Learning, Media and Technology*, vol. 31, no. 3, pp. 233–247, 2006.
- [5] Y. Akpınar and H. Simsek, "Should K-12 Teachers Develop Learning Objects? Evidence from the Field with K-12 Students," *International Journal of Instructional Technology and Distance Learning*, 2007.
- [6] R. Kay, "Examining the Effectiveness of Web-Based Learning Tools in Middle and Secondary School Science Classrooms," *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 7, 2011.
- [7] R. Kay, "Exploring the use of web-based learning tools in secondary school classrooms," *Interactive Learning Environments*, Jan. 2012.
- [8] R. Kay, L. Knaack, and D. Petrarca, "Exploring Teachers Perceptions of Web-Based Learning Tools," *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 5, 2009.
- [9] J. Najjar, J. Klerkx, R. Vuorikari, and E. Duval, "Finding Appropriate Learning Objects: An Empirical Evaluation," in *Research and Advanced Technology for Digital Libraries*, Springer Berlin Heidelberg, 2005, pp. 323–335.
- [10] X. Ochoa and E. Duval, "Relevance Ranking Metrics for Learning Objects," *IEEE Transactions on Learning Technologies*, vol. 1, no. 1, pp. 34–48, 2008.
- [11] T. L. Leacock and J. C. Nesbit, "A Framework for Evaluating the Quality of Multimedia Learning Resources," *Educational Technology and Society*, vol. 10, pp. 44–59, 2007.
- [12] Y. Toll and R. Yohandri, "Aspects and indicators for assessing the quality of learning objects created by the University of Information Sciences, Havana," *Universities and Knowledge Society Journal*, vol. 10, no. 2, pp. 394–406, 2013.
- [13] J. Sanz-Rodríguez, J. M. Doderó, and S. Sánchez-Alonso, "Ranking Learning Objects through Integration of Different Quality Indicators," *IEEE Transactions on Learning Technologies*, vol. 3, no. 4, pp. 358–363, 2010.
- [14] R. Vuorikari, N. Manouselis, and E. Duval, "Using Metadata for Storing, Sharing and Reusing Evaluations for Social Recommendations: the Case of Learning Resources," in *Social information retrieval systems: Emerging technologies and applications for searching the web effectively*, 2008, pp. 87–107.
- [15] X. Ochoa and E. Duval, "Quality Metrics for Learning Object Metadata," in *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2006.
- [16] X. Ochoa and E. Duval, "Towards Automatic Evaluation of Learning Object Metadata Quality," in *Advances in Conceptual Modeling-Theory and Practice*, Springer Berlin Heidelberg, 2006, pp. 372–381.
- [17] J. Sanz-Rodríguez, J. M. Doderó, and S. Sánchez-Alonso, "Metrics-based evaluation of learning object reusability," *Software Quality Journal*, vol. 19, no. 1, pp. 121–140, 2011.
- [18] E. Kurilovas, V. Birenienė, and S. Serikoviene, "Methodology for Evaluating Quality and Reusability of Learning Objects," *Electronic Journal of e-Learning*, vol. 9, no. 1, pp. 39–51, 2011.
- [19] R. H. Kay and L. Knaack, "A multi-component model for assessing learning objects: The learning object evaluation metric (LOEM)," *Australasian Journal of Educational Technology*, vol. 24, no. 5, pp. 574–591, 2008.
- [20] "MERLOT (Multimedia Educational Resource for Learning and Online Teaching)." [Online]. Available: <http://www.merlot.org>.
- [21] R. Kay, "Evaluating learning, design, and engagement in web-based learning tools (WBLTs): The WBLT Evaluation Scale," *Computers in Human Behavior*, vol. 27, no. 5, pp. 1849–1856, Sep. 2011.
- [22] M. Haughey and B. Muirhead, "Evaluating learning objects for schools," *E-Journal of Instructional Sciences and Technology*, vol. 8, no. 1, 2005.
- [23] Y. Eguigure, A. Zapata, V. Menendez, and M. Prieto, "Quality evaluation model for learning objects from pedagogical perspective. A case of study," *Iberoamerican Journal of Applied Computing*, vol. 1, no. 2, 2011.
- [24] X. Ochoa, "Learnometrics: Metrics for Learning Objects," Katholieke Universiteit Leuven, 2008.
- [25] N. Y. Yen, T. K. Shih, L. R. Chao, and Q. Jin, "Ranking Metrics and Search Guidance for Learning Object Repository," *IEEE Transactions on Learning Technologies*, vol. 3, no. 3, pp. 250–264, 2010.
- [26] J. C. Nesbit and T. L. Leacock, "Web-Based Tools for Collaborative Evaluation of Learning Resources," *Journal on Systemics, Cybernetics and Informatics*, vol. 3, no. 5, pp. 102–112, 2006.
- [27] J. Nesbit, K. Belfer, and J. Vargo, "A convergent participation model for evaluation of learning objects," *Canadian Journal of Learning and Technology*, vol. 28, no. 3, 2002.
- [28] J. Vargo, J. C. Nesbit, K. Belfer, and A. Archambault, "Learning object evaluation: computer-mediated collaboration and inter-rater reliability," *International Journal of Computers and Applications*, vol. 25, no. 3, 2003.
- [29] F. Krauss and M. Ally, "A Study of the Design and Evaluation of a Learning Object and Implications for Content Development," *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 1, pp. 1–22, 2005.
- [30] Y. Akpınar, "Validation of a Learning Object Review Instrument: Relationship between Ratings of Learning Objects and Actual Learning Outcomes," *Interdisciplinary Journal of E-Learning and Learning Objects*, vol. 4, 2008.
- [31] R. Luke, D. Mallory, R. Pinet, and A. Segiun, "Evaluating Learning Objects with an Online Version of the Learning Object Review Instrument: Results of a Design Study," in *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 2007.
- [32] A. Gordillo, E. Barra, and J. Quemada, "A flexible open source web platform to facilitate Learning Object evaluation," in *Proceedings of the 2014 Frontiers in Education Conference (FIE 2014)*, 2014.
- [33] E. Barra, A. Gordillo, and J. Quemada, "Virtual Science Hub: An Open Source Platform To Enrich Science Teaching," in *Proceedings of the International Conference on Educational Sciences and Technology (ICEST 2014)*, 2014.
- [34] A. Gordillo, E. Barra, D. Gallego, and J. Quemada, "An online e-Learning authoring tool to create interactive multi-device learning objects using e-Infrastructure resources," in *Proceedings of the 2013 Frontiers in Education Conference (FIE 2013)*, 2013, pp. 1914–1920.
- [35] M. Kendall, *Rank correlation methods*. 1948.
- [36] P. Wessa, "Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7," 2014. [Online]. Available: <http://www.wessa.net>.