Analyzing Associations between the Different Ratings Dimensions of the MERLOT Repository

Cristian Cechinel
Federal University of Pampa,
Bagé, Brazil

Salvador Sánchez-Alonso
University of Alcalá,
Alcalá de Henares, Spain

contato@cristiancechinel.pro.br

salvador.sanchez@uah.es

Abstract

As the dissemination of digital learning resources is continuously growing over the internet, the existing repositories are enabled to search for different alternatives to assess the quality of their materials. MERLOT, one of the most recognized learning object repositories available nowadays, has adopted the peer-review approach as the cornerstone for the quality evaluation of their learning objects. In that evaluation, experts on specific areas rate resources according to three predefined dimensions, and, after an extensive edition process, these ratings are published and used for the recommendation of materials among the community of users. In addition, MERLOT allows users to write comments and provide ratings for the learning resources, complementing its evaluation strategy with this more informal mechanism. The present work analyzes associations between the ratings given by the users and the experts with the aim of discovering whether or not these two groups have similar impressions about the quality of the materials, as well as to explore the usefulness of this twofold strategy towards the establishment of learning resources quality inside MERLOT.

Keywords: Learning objects, quality, ratings, MERLOT, peer-review, public-review

Introduction

As the dissemination and availability of learning resources grows on the Internet, the many existing repositories are searching for mechanisms to evaluate their catalogued/stored materials in order to rank them and better serve the resource seeking needs of their users. Learning object repositories (LORs) are potential aggregators of communities of practitioners (Brosnan, 2005; Han, Kortemeyer, Kramer, & Prummer, 2008; Monge, Ovelar, & Azpeitia, 2008), i.e., people who share interests and concerns about something they do and learn through their interactions (Wenger, 2006). Subsequently, some of the repositories 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 on the evaluations given by regular users and experts that are members of the

repository community. These are forms of evaluative metadata (Vuorikari, Manouselis, & Duval, 2008) that serve as the basis for properly ranking and recommending resources for users.

However, as LORs are distinct in many aspects (number of resources, granularity of the materials, locality of the learning objects, openness for users, subject areas, internal goals, audience)

(McGreal, 2008), so too are the solutions they implement regarding this kind of community-based assessment. For instance, in eLera repository (E-Learning Research and Assessment Network - http://www.elera.net), members can create reviews of the learning objects using the Learning Object Review Instrument (LORI) (Nesbit, Belfer, & Leacock, 2003), and experienced members can moderate teams of members in a collaborative online review process where reviewers discuss and compare their evaluations (Nesbit & Li, 2004). Besides, members can also add some resource to their personal bookmarks, allowing eLera to recommend materials, not only using their associated ratings, but also using their *popularity*. From a different perspective, the Connexions repository (http://cnx.org) approaches quality by a system called *lenses* which arranges resources according to evaluations provided by individuals and organizations (Kelty, Burrus, & Baraniuk, 2008). In this context, resources are explicitly endorsed by third parties and gain higher quality assurance as they start to accumulate more endorsements. Moreover, Connexions also provides mechanisms to sort materials considering their number of accesses over the time and considering the ratings given by the users. The MERLOT repository (Multimedia Educational Resource for Learning and Online Teaching - http://www.merlot.org) introduced a post-publication peer-review model in order to assure the quality of its catalogued resources (Cafolla, 2002). In MERLOT, after their publication, materials are peer-reviewed and rated by experts in the resource domain. As it is not possible to peer-review all resources, MERLOT also allows the community of users to provide their own comments and ratings for the materials, as well as to bookmark their favorite resources in the so-called *Personal Collections*. All this information is used by MERLOT to sort the materials during the searching process.

The case of MERLOT is particularly unusual in the sense that ratings are gathered from two well-defined and distinct groups, public and experts (it becomes important to mention here that peer-reviewers in MERLOT are also members which may cause some overlap of individuals in these two groups of people), which possibly come from distinct backgrounds and may have divergent opinions with respect to quality. In fact, the differences between these groups could be considered the strength of the adopted approach, once it provides complementary views about the same subject. Considering this, the present paper analyzes the existence of associations between the ratings given by these two groups of evaluators in MERLOT in order to discover whether or not they diverge about the quality assessment of the same materials, as well as to explore the usefulness of such complementary evaluations towards the assurance of quality inside the repository.

The rest of this paper is structured as follows. The next section describes the features and differences between peer-review and public-review systems, and the third section presents how these systems are applied in MERLOT. The fourth section reports and discusses the exploratory analysis performed to evaluate the existence of associations between the ratings given by the different groups of evaluators in MERLOT. Finally, the fifth section presents conclusions and future work.

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, who requests a group of fellow-experts to review it. The review process provides advice about whether or not the article can be accepted for publishing, and what further work is still required in the case of acceptance (Harnad, 2000). 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 published papers. However, the system is not free from criticisms, and issues such as conflicts of interest, biases of the peers, time delays, and the difficulty to detecting fraud are often mentioned as possible shortcomings of the peer-review

process (Benos et al., 2007). Despite the controversies regarding its efficiency, the peer-review system remains as the cornerstone of quality assurance in the academic field and has also entered in the scene of educational resources after its implementation in MERLOT.

Contrary to peer-review systems, mainly related to the scientific field, public-review is widely diffused in many other areas, such as online vendors (e.g., Amazon - http://www.amazon.com, eBay - http://www.ebay.com) and several communities of interest (e.g., IMDb - http://www.imdb.com, YouTube - http://www.youtube.com, RYM - http://rateyourmusic.com). In these, users normally benefit from comments and ratings given by the community through the use of recommender systems (such as collaborative filters) which, based on the comparison of users' profiles and the correlation of personal tastes, provide personalized recommendation of items and products that are suggested to be of the users' interest (Resnick & Varian, 1997). In this kind of social system, 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 (Peddibhotla & Subramani, 2007). Table 1 explores some other aspects which normally differentiate standard peer-review and public-review systems.

Table 1: Different aspects involving peer-review and public-review

ASPECTS	PEER-REVIEW	PUBLIC-REVIEW
Evaluator Background	Expert in the field domain	Non-expert
Existence of official criteria or metrics	Yes	No/Sometimes
Size of the community of evaluators	Restricted	Wide opened
Common Models	Pre-publication	Post-publication
Domain	Scientific field, journals and funding calls	Online vendors, communities of interest
Motivation	Prestige, fame, to determine the quality and direction of research in a particular domain, obligation	Desire and need of social interaction, professional self expression, reputation
Communication among evaluators	Not allowed	Encouraged
Selection of evaluators	Editor Responsibility	None
Financial Compensation	Normally none	None
Time taken for the evaluation	Typically Slow	Typically Fast
Level of formality	Formal process for editing and revision	Informal
Author's identity	Masked	Non-masked
Requirements to be a reviewer	To be an expert in the field and to be invited	Creation of a member's account

Reviews and Ratings in MERLOT

The Multimedia Educational Resource for Learning and Online Teaching (MERLOT) is an international initiative that allows users to catalogue educational resources with the aim of facilitating the use and sharing of online learning technologies. MERLOT has adopted a post publication peer-review model (Cafolla, 2002), where already catalogued materials are peer-reviewed by discipline experts who are members of some discipline community editorial board (e.g., Biology,

Business, Chemistry, Mathematics, Psychology). Editorial boards of MERLOT decide on the process of selecting materials that are worthy of reviewing, and the assigned materials are then independently peer-reviewed by their members according to three main criteria: 1) *Quality of Content*, 2) *Potential Effectiveness as a Teaching Tool*, and 3) *Ease of use*. After peer-reviewers report their evaluations, the editorial board chief-editor composes a one single report and publishes it in the repository with the permission of the authors (Merlot, 2010a). In MERLOT, the formation of the editorial boards is mostly through the indication of individuals by MERLOT institution partners, but also by volunteers who meet some minimum criteria (e.g., being an instructor in an institution, being an expert in the scholarship of their field, possessing excellence in teaching) and participate in the training offered by MERLOT.

In addition to peer-review evaluations, MERLOT also allows the registered members of the community to provide comments and ratings about the materials, complementing its strategy of evaluation with an alternative and more informal mechanism. Besides the ratings checkbox and the form where the member can fill in his or her comments, the MERLOT's *Add Comment Section* also asks if the member has used the resource in the classroom and offers two extra forms where the members can make technical remarks and inform the time they spent reviewing the material. The ratings of users and peer-reviewers range from 1 to 5 (with 5 as the best rating). The use of the same rating scales for both kinds of evaluations allows us to contrast these ratings in order to evaluate possible correlations and the existence or not of disagreement between the two groups of evaluators.

Exploratory Analysis

Data Sample

The method followed for the study reported here was the analysis of correlations between the ratings given by the two kinds of evaluators in MERLOT, named as peer-reviewer's ratings (PRR) and user's ratings (UR). Data from a total of 20,506 learning objects was gathered (September 2009) through a web crawler developed for that purpose and similar in functionality to the one reported by Biletskiy, Wojcenovic, and Baghi (2009). Most of the resources did not already have any instances of peer-review or user rating, and, from the collected data, only 3.38% presented both at least one peer-review and at least one user rating (Table 2).

Table 2. Sample sizes of peer-reviewed and user-reviewed resources

TOTAL SAMPLE SIZE	PRR > 0		UR > 0		PRR ∩ UR	
	SIZE	%	SIZE	%	SIZE	%
20,506	2,595	12.65	2510	12.24	695	3.38

This sample containing ratings from both groups at the same time (PRR ∩ UR) was used in the present study and the rest of the data was discarded.

Analyzing the Associations

Users ratings against overall peer-reviewers ratings

As both samples did not follow a normal distribution, a non-parametric analysis was performed using the Spearman's rank correlation (r_s) to evaluate whether or not there is association between the ratings of the two groups, i.e., if the raters agree or not about the quality of the resources. Considering that MERLOT divides resources into different categories of disciplines, we split the

collected sample according to these categories. This allows us to observe potential differences according to the background of the evaluators. As MERLOT allows users to catalogue the materials in more than one category of discipline, some part of the split sample is overlapped. However, we decided to maintain the learning objects classified in more than one discipline due to the fact we considered this overlap relatively small (16%). Table 3 presents the results of this analysis.

Table 3. Correlation between users ratings and peer-reviewers ratings considering the categories of disciplines

Discipline	Sample Size	PRR Average (std)	UR Average (std)	r_S	P	S
All	695	4.34(0.70)	4.29(0.70)	0.19	0.00	Y
Arts	25	4.14(0.74)	4.43(0.58)	0.20	0.33	N
Business	59	4.22(0.79)	4.15(0.94)	0.06	0.66	N
Education	167	4.41(0.68)	4.36(0.72)	0.16	0.04	Y*
Humanities	133	4.60(0.51)	4.40(0.67)	0.19	0.03	Y
Mathematics & Statistics	66	4.67(0.52)	4.25(0.69)	0.17	0.31	N
Science & Technology	285	4.21(0.71)	4.25(0.72)	0.26	0.00	Y
Social Sciences	73	4.20(0.75)	4.38(0.60)	0.2	0.09	Y$^+$

In Table 3, the column P stands for the p-value obtained in the analysis. In the column S, N stands for no significant association between the ratings given by the two groups of evaluators, Y represents significant association at 99% level, Y* means significant association at 95% level, and Y$^+$ at 90% level. The correlation coefficient (r_s) indicates the strength of the association between the two groups of ratings varying from -1 to 1, where 0 means no association (no agreement). The closer the correlation coefficient is to 1, the better is the association. As can be seen in Table 3, the disciplines of *Arts*, *Business*, and *Mathematics and Statistics* did not present any association between the ratings given by users and peer-reviewers. However, the ratings for the disciplines of *Education*, *Humanities*, *Science and Technology* and *Social Sciences*, and for the overall sample did presented an association. But, even though these associations exist, they are not too strong, as their coefficients of correlation are relatively small. Figure 1 better illustrates the weakness of the association for the discipline of *Science and Technology* (we selected this discipline due to the fact that this was the one which has presented the highest coefficient of correlation of all).

As can be seen in Figure 1, it is not possible to observe patterns indicating concordance between the ratings of the two groups. (A strong correlation between the ratings could be suggested by a formation of a diagonal line, or the agglomeration of dots in some region of the matrix, for instance). In fact, we can observe several cases where users and peer-reviewers strongly disagree about the ratings. The weakness of the associations is also confirmed when we perform a linear regression analysis in order to explore more deeply the relationships between users and peer-reviewers ratings in the discipline of *Science and Technology*. Despite the fact that it is possible to generate a linear prediction model at a 99% level of significance, the coefficient of correlation remains small (0.28) and the model is able to represent only 7.94% of the entire population.
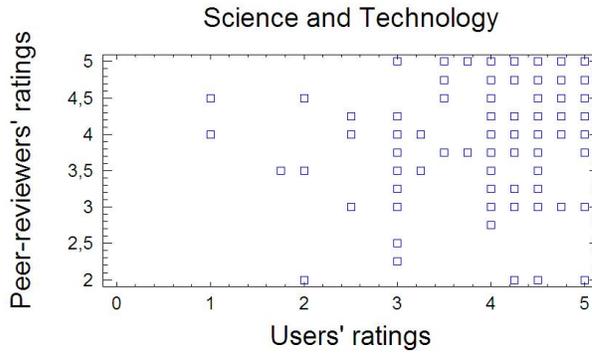
Figure 1: Scatter plot matrix of ratings for the discipline of *Science and Technology*

At first glance, this exploratory analysis indicates that both groups of reviewers have different impressions about the quality of the learning objects catalogued in MERLOT, thus serving as complementary views of assessment inside the repository.

Users ratings against the three criteria of peer-reviewers ratings

As mentioned before, the peer-reviewer rating is composed of three main criteria: 1) *Quality of Content*, 2) *Potential Effectiveness as a Teaching Tool*, and 3) *Ease of use*. As the existence (or the absence) of associations could be specifically related to one of these dimensions, we ran the same analysis to evaluate the associations between the users' ratings and each one of the criteria of peer-reviewers' ratings. Table 4 presents the results of this analysis.

**Table 4. Correlation between users ratings and the three criteria
of peer-reviewers ratings**

Discipline	Quality of content			Potential effective as a teaching tool			Ease of Use			Total Rating
	r_S	P	S	r_S	P	S	r_S	P	S	S
All	0.21	0.00	Y	0.25	0.00	Y	0.19	0.00	Y	Y
Arts	0.21	0.30	N	0.14	0.50	N	0.41	0.04	Y*	N
Business	0.07	0.61	N	0.04	0.73	N	0.13	0.32	N	N
Education	0.11	0.15	N	0.30	0.00	Y	0.15	0.05	Y*	Y*
Humanities	0.27	0.01	Y	0.26	0.00	Y	0.15	0.09	Y+	Y
Mathematics & Statistics	0.23	0.06	Y*	0.02	0.87	N	0.02	0.85	N	N
Science & Technology	0.25	0.00	Y	0.32	0.00	Y	0.25	0.00	Y	Y
Social Sciences	0.25	0.03	Y*	0.27	0.05	Y*	0.11	0.33	N	Y+

From the table it can be seen that in some disciplines the associations encountered before do not persist for all three evaluation criteria. For instance, in the *Education* discipline, users' ratings are associated with only two of the three evaluation criteria, more precisely *Potential effectiveness as a teaching tool* and *Ease of use,* and in the discipline of *Social Sciences* users' ratings are associated only to the *Quality of content* and *Potential effectiveness as a teaching tool* criteria. Moreover, other disciplines that did not present any association between the two groups of ratings now present association between users' ratings and some specific evaluation criteria. This is the case of the discipline of *Mathematics and Statistics*, which presented association between the users'

ratings and the *Quality of content* criteria, and of the discipline of *Arts*, which presented association between users' ratings and the *Ease of use* criteria.

It is interesting to highlight that the associations encountered in this analysis are slightly stronger than the associations encountered before (with just two exceptions). However, the coefficients of correlations found are still weak, which reinforces the initial conclusion that users and peer-reviewers have different impressions about quality. Such weakness is again confirmed when we perform a linear regression analysis between users ratings and the *Ease of use* criteria in the discipline of *Arts* (these are the data sets which presented the highest coefficient of correlation, with $r_s = 0.45$). Here, the prediction model is generated at a 95% level of significance, however, the coefficient of correlation is small (0.46) and the model represents only 21% of the entire population.

As the evaluations given by users do not follow any pre-defined criteria, it is difficult to precisely understand what they are referring to without an in-depth look at the textual information (comments) attached to them. Bearing this in mind, one of the two following situations may be occurring: 1) the impressions of quality that users have are not related to any one of the criteria evaluated by peer-reviewers; or 2) if users evaluate the same aspects as peer-reviewers, they do not agree with the ratings given by the experts about these aspects. From our point of view, these two situations alternate depending on the category of the discipline and on the peer-review criteria under evaluation. Moreover, the fact that some associations observed in the first analysis now present a stronger coefficient of correlation in this second analysis may be an indication that, for these cases, the situation number 2 is occurring. However, these are all assumptions that still require more investigation to be confirmed.

Conclusions and Future Work

The most important contribution of this paper is the indication that both communities of evaluators in MERLOT are communicating different views regarding the quality of the learning objects refereed in the repository. Even though we have found associations between the users' ratings and the peer-reviewers' ratings in some disciplines, such associations are relatively weak and can not confirm that users and experts agree about the quality of the evaluated learning resources. This reinforces the idea that peer-review and public-review approaches can be adopted in learning objects repositories as complementary strategies of evaluation that can both serve for the assurance and the establishment of quality parameters for further recommendation of materials.

MERLOT has been successfully used and recognized mostly because of the implementation of the peer-review system, which still remains its cornerstone strategy for quality evaluation. However, the community of members and their ratings in MERLOT are naturally growing much more than the community of peer-reviewers and their evaluations. During the 30 days of preparation of this paper the number of new members has increased by 1,426, whereas the number of peer-reviewers by 5. Moreover, the number of new comments was approximately 6.5 times higher than the number of new peer-reviews (MERLOT, 2010b). Such rapid increases makes it necessary to invest attention in exploring the inherent potentialities of this expanding community, as it was done for instance by Sicilia, Sanchez-Alonso, García-Barriocanal, & Rodríguez-García (2009). Future work will qualitatively examine existing divergences between these two kinds of evaluations, as well as explore the ratings of groups of learning objects that are classified in more than one category and evaluate the utilization of collaborative filtering in the recommendation of learning objects inside MERLOT.

References

Benos, D. J., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., . . . Zotov, A. (2007). The ups and downs of peer review. *Advances in Physiology Education, 31*, 145-152.

Biletskiy, Y., Wojcenovic, M., & Baghi, H. (2009). Focused crawling for downloading learning objects - An architectural perspective. *Interdisciplinary Journal of E-Learning and Learning Objects*, 5, 169-180. Retrieved from http://www.ijello.org/Volume5/IJELLOv5p169-180Biletskiy416.pdf

Brosnan, K. (2005). Developing and sustaining a national learning-object sharing network: A social capital theory perspective. In J. B. Williams, & M. A. Goldberg (Eds.), *Proceedings of The ASCILITE 2005 Conference*, 105-114. Brisbane, Australia.

Cafolla, R. (2002). Project Merlot: Bringing peer review to web-based educational resources. *Proceedings of the USA Society for Information Technology and Teacher Education International Conference*, 614–618.

Han, P., Kortemeyer, G., Kramer, B. J., & Prummer, C. (2008). Exposure and support of latent social networks among learning object repository users. *Journal of Universal Computer Science (JUCS)*, *14*(10), 1717-1738.

Harnad, S. (2000). The invisible hand of peer review. *Exploit Interactive*, 5(April). Retrieved from http://www.exploit-lib.org/issue5/peer-review

Kelty, C. M., Burrus, C. S., & Baraniuk, R. G. (2008). Peer review anew: Three principles and a case study in postpublication quality assurance. *Proceedings of the IEEE, 96*(6), 1000–1011.

McGreal, R. (2008). A typology of learning object repositories. In H. H. Adelsberger, Kinshuk, J. M. Pawlowski, & D. G. Sampson (Eds.), *International handbooks on information systems: Handbook on information technologies for education and training* (pp. 5-28). Heidelberg: Springer.

Merlot. (2010a). *About us: MERLOT Peer Review Process*. Retrieved from http://taste.merlot.org/peerreviewprocess.html

Merlot. (2010b). *What's New*. Retrieved from http://www.merlot.org/merlot/whatsNew.htm

Monge, S., Ovelar R., & Azpeitia, I. (2008). Repository 2.0: Social dynamics to support community building in learning object repositories. *Interdisciplinary Journal of E-Learning and Learning Objects, 4,* 191-204. Retrieved from http://www.ijello.org/Volume4/IJELLOv4p191-204Monge.pdf

Nesbit, J. C., & Li, J. (2004). Web-based tools for learning object evaluation. *Proceeding of the International Conference on Education and Information Systems: Technologies and Applications*. Orlando, Florida.

Nesbit, J. C., Belfer, K., & Leacock, T. (2003). Learning Object Review Instrument (LORI). *E-Learning Research and Assessment Network*. Retrieved from http://www.elera.net/eLera/Home/Articles/LORI%201.5.pdf

Peddibhotla, N., & Subramani, M. R. (2007). Contributing to public document repositories: A critical mass theory perspective. *Organizational Studies, 28*(3), 327-346.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of ACM*, *40*(3), 56-58.

Sicilia, M-A., Sanchez-Alonso, S., García-Barriocanal, E., & Rodríguez-García, D. (2009). Exploring structural prestige in learning object repositories: Some insights from examining references in MERLOT. *Proceedings of the International Conference on Intelligent Networking and Collaborative Systems*, 212-218.

Vuorikari, R., Manouselis, N., & Duval, E. (2008). Using metadata for storing, sharing and reusing evaluations for social recommendations: The case of learning resources. In D. Goh & S. Foo (Eds.), *Social information retrieval systems: Emerging technologies and applications for searching the web effectively* (pp. 87-107). New York, NY: Idea Group.

Wenger, E. (2006). *Communities of practice: A brief introduction.* Retrieved from
   http://www.ewenger.com/theory/communities_of_practice_intro.htm

Biographies

M.Sc. **Cristian Cechinel** is a professor at the Computer Engineering Course of Federal University of Pampa. He obtained his Bachelor's and Master's degree in Computer Science from Federal University of Santa Catarina, and he is currently taking his Ph.D. on Learning objects quality at the Computer Science Department of University of Alcalá. His research focuses on Learning Technologies and Artificial Intelligence.

Dr. Salvador Sanchez-Alonso is an associate professor in the Computer Science Department of the University of Alcalá and a senior member of the Information Engineering research unit of the same university. He previously worked as an assistant professor at the Pontifical University of Salamanca for 7 years during different periods, and also as a software engineer at a software solutions company during 2000 and 2001. He earned a Ph.D. in Computer Science at the Polytechnic University of Madrid in 2005 with a research on learning object metadata design for better machine "understandability". His current research interests include Learning technologies, Semantic Web and Computer science education.