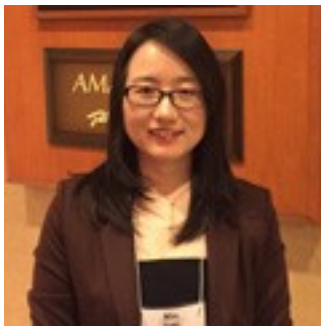# Not All Rubrics Are Equal: A Review of Rubrics for Evaluating the Quality of Open Educational Resources

Min Yuan, and Mimi Recker
Utah State University, USA

Abstract

The rapid growth in Internet technologies has led to a proliferation in the number of Open Educational Resources (OER), making the evaluation of OER quality a pressing need. In response, a number of rubrics have been developed to help guide the evaluation of OER quality; these, however, have had little accompanying evaluation of their utility or usability. This article presents a systematic review of 14 existing quality rubrics developed for OER evaluation. These quality rubrics are described and compared in terms of content, development processes, and application contexts, as well as, the kind of support they provide for users. Results from this research reveal a great diversity between these rubrics, providing users with a wide variety of options. Moreover, the widespread lack of rating scales, scoring guides, empirical testing, and iterative revisions for many of these rubrics raises reliability and validity concerns. Finally, rubrics implement varying amounts of user support, affecting their overall usability and educational utility.

Keyword: Rubrics, Quality, Open Educational Resources, Content, Development Process, Application Context, Support

Introduction

Open Educational Resources (OER) are online teaching, learning, and research resources that can be freely accessed, adapted, used, and shared to support education (U.S. DoE, 2010). Fueled in part by the rapid growth in Internet technologies, a broad range of OER has become widely availability, providing a content infrastructure with the potential for greatly enhancing teaching and learning (Atkins, Brown, & Hammond, 2007; Borgman et al., 2008; Porcello & Hsi, 2013). For example, OER can support teachers in gaining and sharing content and pedagogical knowledge, and can provide learners access to a wide variety of resources for extending their knowledge and skills (Haughey & Muirhead, 2005; Kay & Knaack, 2007; Khanna & Basak, 2013). In addition, OER can radically change the way information is presented and the way learners engage with information. Some OER contain images, videos, or interactive content, for instance, helping to make abstract concepts more concrete, while other OER can be adapted to fit learners' different needs (Kay & Knaack, 2007).

Nonetheless, the wide availability of OER does not assure their high quality or educational utility (Porcello & Hsi, 2013; Rodríguez, Dodero, & Alonso, 2011), as both high-quality and low-quality OER can be found throughout the Internet (Bundsgaard & Hansen, 2011; Fitzgerald, Lovin, & Branch, 2003). Moreover, low-quality OER can hamper instructional practices and waste teachers' limited time (Abramovich, Schunn, & Correnti, 2013; Wetzler et al., 2013). As a result, evaluating the quality and appropriateness of OER has become a pressing need (Porcello & Hsi, 2013).

In response, several researchers and educational organizations have been developing rubrics to help guide the judgment of OER quality. As described below, these rubrics vary widely along a number of critical dimensions. For example, the Learning Object Review Instrument (LORI) from Nesbit, Belfer, and Leacock (2007) is designed to evaluate a wide variety of OER, while the Learning Object Evaluation Instrument (LOEI) from Haughey and Muirhead (2005) is designed for school contexts. The Educators Evaluating the Quality of Instructional Products (EQuIP) from Achieve (2014) focuses on the alignment of OER with educational standards, while the rubric from Fitzgerald and Byers (2002) is targeted at inquiry-based science resources. As a final example, the Achieve organization, which developed the rubric for Evaluating Open Education Resource Objects (OER rubric) in 2011 and the EQuIP rubric in 2014, provides extensive training materials for users of these two rubrics, while developers of some other rubrics do not.

In sum, different rubrics possess different characteristics and emphasize different aspects, which can lead to confusion when deciding which rubric to use for OER evaluation. Therefore, in an attempt to synthesize the state of the field, this article provides a review of existing quality rubrics for OER evaluation, and compares them along key characteristics and the kinds of support provided to users.

Theoretical Context

Defining Rubrics

A rubric provides a scoring scheme to help guide a user in judging products or activities (Moskal, 2000). For example, rubrics are widely used in education to help guide people's evaluation of a variety of constructs, including students' writing performances, the quality of research projects, and the quality of educational resources (Bresciani et al., 2009; Custard & Sumner, 2005; Rezaei & Lovorn, 2010). Among them, a number of rubrics have been developed to evaluate the quality of OER, as people increasingly need assistance in identifying high-quality resources available on the Internet (e.g., Custard & Sumner, 2005; Haughey & Muirhead, 2005; Porcello & Hsi, 2013).

However, the use of rubrics does not always lead to improved evaluation (Rezaei & Lovorn, 2010). Indeed, an important consideration is a rubric's validity and reliability (e.g. Bresciani, et al., 2009; Jonsson & Svingby, 2007; Rezaei & Lovorn, 2010). Validity is the extent to which a rubric measures what it is purported to measure, while reliability is the extent to which the results from a rubric are consistent over time and across different raters (Kimberlin & Winterstein, 2008; Moskal & Leydens, 2000). Further, utility is another measure of the quality of a rubric, and has been promoted by many researchers (e.g., Ross, 2006; Willner, 1997).

Researchers and developers have taken different approaches to improving the performance of rubrics, such as evaluating the validity and reliability of rubrics through empirical testing, and improving the utility of rubrics by providing user support (Colton et al., 1997; Moskal & Leydens, 2000; Wolfe, Kao, & Ranney, 1998). Thus, to better understand the performance of different rubrics for evaluating the quality of OER, a review and synthesis of these rubrics along these different dimensions is needed.

Characterizing Rubrics

A rubric typically focuses on specific *content*, follows a particular *development process*, and targets at a particular *application context* (e.g., Arter & Chappuis, 2006; Moskal, 2000; Moskal & Leydens, 2000). Thus, we analyze OER quality rubrics following these three aspects.

The *content* aspect focuses on how the rubric deconstructs overall OER quality. In some rubrics, OER quality is defined in terms of multiple dimensions, where, in turn, each quality dimension can be comprised of one or more quality indicators. For example, the rubric from Pérez-Mateo, Maina, Guitert, and Romero (2011) first identifies three dimensions of OER quality – content, format, and process, and then proposes 42 indicators for these three dimensions. In contrast, other rubrics only identify quality indicators (e.g., the LORI). The *content* aspect also addresses whether rubrics have a ratings scale, and/or provide a detailed scoring guide.

Second, the *rubric development processes* examines whether rubrics have reported empirical testing results, and whether they have been iteratively improved. Third, the *application context* aspect examines whether rubrics apply to a variety of OER, or are specific to a particular website or discipline, and whether the rubrics were designed for human or machine use.

Exploring Support for Rubric Users

Simply creating a rubric is not sufficient. A rubric is also expected to be usable and to improve evaluation. The utility and effectiveness of a rubric depends on a variety of factors. For example, Colton et al. (1997) and Wolfe et al. (1998) noted that appropriate training about features of the rubric and strategies for using a rubric can improve its validity and reliability. Similarly, Rezaei and Lovorn (2010) argued that without appropriate user support, the use of rubrics may not necessarily improve the reliability or validity of assessment.

Methods

We conducted a search for rubrics designed to evaluate OER over a six-month period, ending in April 2014. In particular, we searched ERIC, Education Full Text, Library Literature and Information Sciences, and Google Scholar, using different combinations of the following descriptors: quality, open educational resources (OER), rubrics, evaluation, judgment, and assessment. We also identified articles by reviewing the reference lists of existing rubrics, and getting recommendations from other researchers. These different strategies led to more than 200 articles. However, as many resulting articles did not actually propose a rubric and some resulting rubrics were not designed for evaluating OER quality, we established the following inclusion criteria.

To qualify for inclusion in this review, each quality rubric had to meet four criteria: 1) it evaluated *quality* instead of relevance, or other features; 2) it focused on the *OER* instead of associated programs, services, or instruction; 3) it included a concrete rubric; 4) it provided an introduction or an explanation of the rubric; and 5) it was published later than 2000. Fourteen rubrics were ultimately included in our analysis.

The 14 rubrics were analyzed in two stages. The first stage examined rubrics in terms of the *content, development process, application context* framework. The second stage examined how rubrics provided user support.

Findings

Rubric Characteristics

The Appendix shows a summary of the 14 selected rubrics following the three aspects described above, and Table 1 shows frequencies of the various aspects in the selected rubrics. These results are explained in more detail below:

Table 1. Frequency of Different Aspects for Selected Rubrics

Aspect	Category	Frequency	Percent
Content			

Quality Dimension	a. Rubrics identifying both quality dimensions and quality indicators	5	35.7%
	b. Rubrics identifying only quality indicators	9	64.3%
Rating Scale	a. Not specified	5	35.7%
	b. Binary rating scale	2	14.3%
	c. Four-point rating scale	2	14.3%
	d. Five-point rating scale	5	35.7%
Scoring Guide	a. Rubrics provide detailed scoring guide	3	21.4%
	b. Rubrics do not provide detailed scoring guide	11	78.6%

Rubric Development Processes

Tested vs. Not Tested	a. Rubrics did not report empirical testing results	8	57.2%
	b. Rubrics reported empirical testing results	6	42.8%
Revised vs. Not Revised	a. Rubrics revised several times	3	21.4%
	b. Rubrics not revised	11	78.6%

Application Context

Generic vs. Specific	a. Generic rubrics	7	50.0%
	b. Rubrics specific to a website	5	35.7%
	c. Rubrics specific to a discipline	2	14.3%
Automated vs. Manual	a. Rubrics suited to automated evaluation	2	14.3%
	b. Rubrics suited to manual evaluation	12	85.7%

Content. First, while these rubrics all purport to measure OER quality, they emphasize different aspects of this construct. For example, the OER rubric from Achieve (2011) focuses on the pedagogical value of OER, the rubric from Kurilovas, Bireniene, and Serikoviene (2011) highlights the reusability of OER, while the EQuIP rubric from Achieve (2014) emphasizes OER alignment to educational standards.

Second, all rubrics choose to deconstruct the quality construct into many dimensions and/or indicators. Specifically, as shown in Table 1, about one third of the rubrics (36%) deconstruct quality into dimensions (each, in turn, comprised of one or more quality indicators), thereby reflecting a more nuanced notion of quality, while the majority (64%) use only quality indicators. Moreover, while all rubrics are comprised of a series of quality indicators, the number of indicators in each rubric shows a wide spread ($Mean$ = 15.53, Med= 15.00; SD=10.72; Min=3; Max=42). Figure 1 depicts the number of quality indicators across these rubrics. Note that the EQuIP rubric proposes different numbers of quality indicators for different disciplines (see Appendix), and an average is reported in Figure 1. It is also important to note that while a larger number of indicators can provide a more comprehensive evaluation, it comes with an associated burden on users in terms of applying each indicator in the rubric (Haughey & Muirhead, 2005).
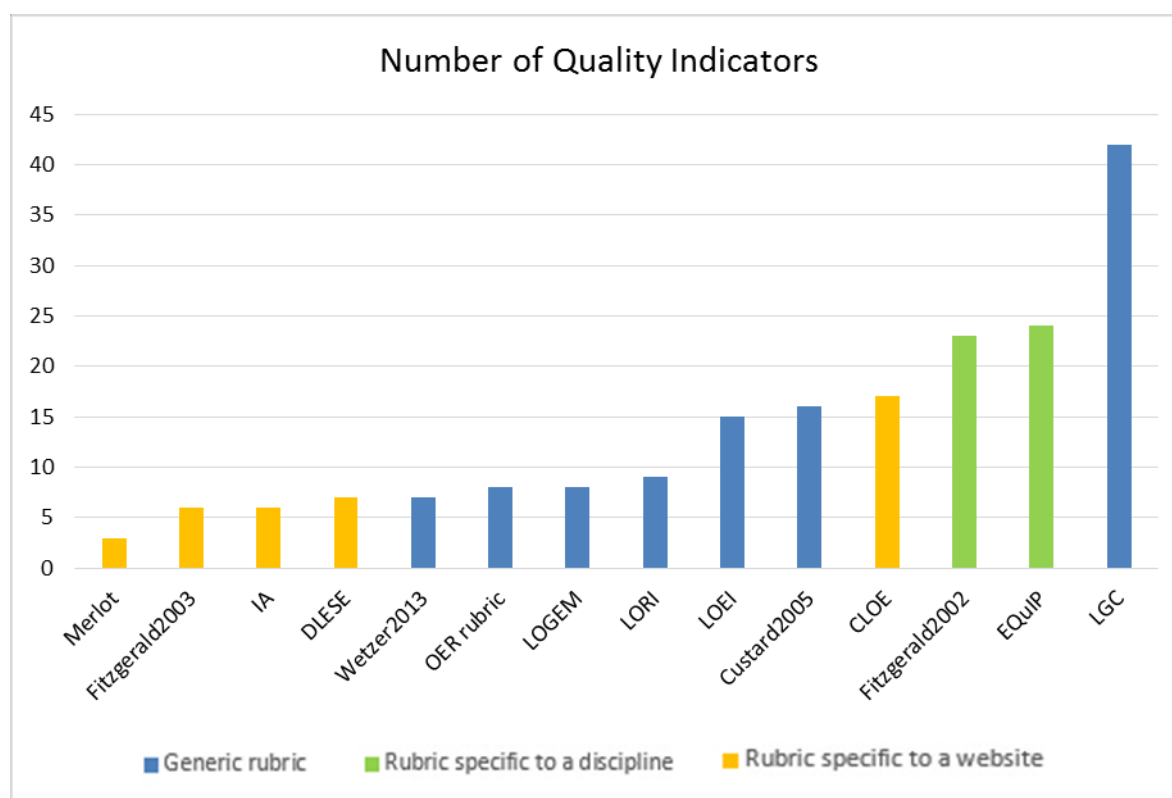

Figure 1. Number of quality indicators across rubrics, characterized by whether it's a generic or a specific rubric

Third, these rubrics share some common indicators. For example, many rubrics include "content quality" as a dimension for evaluating OER. Several rubrics elaborate this construct into indicators such as

completeness, clarity, and accuracy (Fitzgerald et al., 2003; Nesbit et al., 2007; Leary, Giersch, Walker, & Recker, 2009). In contrast, some rubrics also contain unique indicators. For example, the rubric from Haughey and Muirhead (2005) uniquely includes "value" as an important quality indicator of OER, emphasizing that OER should be appropriate for students with diverse needs, languages, and cultures. In addition, the rubric from Wetzler et al. (2013) is the only rubric that considers "sponsor" of OER as an important factor comprising OER quality.

Fourth, the use of rating scales in these rubrics differs. Among the 14 rubrics, five do not specify a rating scale at all. For the remaining nine rubrics, two adopt a binary (yes – no) rating scale. The other seven rubrics adopt a four-point or five-point rating scale, which may provide users with more discriminating power when rating OER.

Fifth, only the OER rubric from Achieve (2011), the EQuIP rubric from Achieve (2014), and the LORI from Nesbit et al. (2007) provide detailed scoring guides. These scoring guides list the steps to carry out the assessment, identify different requirements for different points on the scale, and/or provide examples in order to help users provide more accurate ratings. Note that the provision of rating scales and scoring guides can facilitate users' application of these rubrics and thereby improve the performance (e.g., validity, reliability) of these rubrics (Barkaoui, 2010; Yuan, Recker, & Diekema, , 2015).

Rubric development process. The rubrics were all developed through reviewing existing materials and prior studies, and thus not designed in a void. For example, Vargo, Nesbit, Belfer, and Archambault (2003) design and propose the LORI by reviewing work aimed at evaluating the quality of textbooks, courses, and instructional programs. However, among the 14 rubrics, only six have reported empirical results regarding their validity, reliability, or utility. Such studies can help increase the credibility of these rubrics as well as provide data for iterative improvement of the rubrics.

Finally, three rubrics – the OER rubric from Achieve (2011), the LORI from Nesbit et al. (2007), and the EQuIP rubric from Achieve (2014) – have been revised several times, while the remaining have not.

Application context. The 14 rubrics can be categorized in terms of their application contexts. As shown in Figure 1, seven rubrics (e.g., the OER rubric from Achieve (2011)) are very generic, and thus can be used to evaluate a variety of OER across different resource types and different subject domains. Another five rubrics are specific to particular websites (e.g., the rubric from Leary et al. (2009) applies to the *Instructional Architect* website). Two rubrics are specific to particular subject domains. In particular, the rubric from Fitzgerald and Byers (2002) is specific to the science domain, while the EQuIP rubric from Achieve (2014) is specific to math, literacy, and science.

Another important consideration is the target user: human or machine. Two rubrics – the rubrics from Custard and Sumner (2005) and Wetzler et al. (2013) – are aimed toward automated evaluation of OER quality. The remaining 12 rubrics are designed for use by people.

In summary, despite of some commonalities, these rubrics show a wide diversity in terms of their content, development processes, and application contexts, which give users many choices. However, a lack of rating scales, scoring guides, empirical testing, and iterative revisions for many rubrics calls into question issues surrounding rubric reliability, validity, and utility.

Rubric Support for Users

Our analysis revealed that the rubrics provide a variety of support structures for users. First, many rubrics make themselves easily accessible. For example, after its release at no cost, the LORI was cited by many educational organizations, and consequently used by more teachers (Akpinar, 2008; Nesbit & Li, 2004). Further, the LORI was referred to and studied by many researchers, which in turn lead to suggestions for revisions of this rubric, as well as a series of new rubrics (Haughey & Muirhead, 2005).

Second, many rubric developers solicited user input. For example, the rubric from Pérez-Mateo et al. (2011) asked potential rubric users to suggest a set of indicators that can be used to evaluate OER quality. In particular, based on 114 participants' responses to an online survey, and validated with existing literature, Pérez-Mateo et al. (2011) identified 42 quality indicators, including indicators such as adequacy, consistency, and effectiveness. As another example, the EQuIP rubric from Achieve (2014) asked teachers to submit teaching resources, and a review panel reviewed and provided feedback using their rubric.

Third, some rubrics provide training materials. The Achieve organization provides training materials in PDF, PowerPoint, and video format for both the OER and the EQuIP rubrics. For example, the training materials associated with the OER rubric include a detailed handbook, which introduces the rubric, lists the steps for using the rubric, and provides links to examples. Additionally, rubric developers provide videos explaining the rating scale and showing how to apply rubrics in authentic situations.

Fourth, some rubric developers received support from government and other educational organizations. The Achieve organization, for example, collaborated with several U.S. states (e.g., California, Illinois). In particular, the Achieve organization introduced the basic concepts of OER to different states, developed recommendations for states on the use of rubrics for evaluating OER, helped them develop relevant strategies, and assisted them in implementing these strategies. In this way, the Achieve increased the awareness of OER in these states, which made the use of its rubrics very popular. Finally, developers of the Kurilovas et al. (2011) rubric worked with several European educational organizations, which made their rubric more popular in Europe.

Conclusion

This article first reviewed existing quality rubrics in terms of their content, development processes, and application contexts. For content, even though these rubrics were all comprised of a set of quality indicators and shared similar indicators, different rubrics had different emphases and some contained unique indicators. Over 35% of the reviewed rubrics did not provide a rating scale, and only a small

proportion offered detailed scoring guides. In terms of development processes, over 50 % of the reviewed rubrics had not reported results from empirical testing, and only a small proportion of the rubrics had been revised several times. Lastly, the application contexts of rubrics differed in that some rubrics were more generic than others, and some rubrics were designed to support automated evaluation.

Thus, the research revealed a complex picture. On the positive side, rubrics showed great heterogeneity in the three aspects, providing users with multiple options and for diverse educational applications. On the negative side, only some rubrics provided rating scales, scoring guides, or empirical testing results. As shown in previous studies, the use of rating scales and scoring guides can improve users' evaluation reliability, and empirical testing and the iterative revision of rubrics allowed rubric developers to increasingly improve rubrics' validity, reliability, and usability (Akpinar, 2008; Barkaoui, 2010; Moskal & Leydens, 2000; Vargo et al., 2003). Thus, the absence of rating scales, scoring guides, empirical testing, and iterative revisions in many rubrics raises concerns about their overall utility.

Additionally, this article reviewed the kinds of user support provided by these rubrics, and revealed that these supports came in various forms, such as providing training materials and soliciting user input. These supports were intended to increase the application scope and educational value of these rubrics, and thus are recommended as an important component of future rubric development.

The significance of this research lay in the following aspects. First, it revealed the current state of existing rubrics for evaluating OER quality, showed the characteristics of these rubrics, and identified what supporting structures were provided to users. Thus, it provided a basis for future research on rubrics.

Second, this review indicated that the quality and educational utility of OER depended not only on the content quality (e.g., accuracy, clarity) of resources, but also on the pedagogical values contained in these OER. In particular, these rubrics highlighted some pedagogical guidelines, such as aligning with standards, identifying appropriate learners, and showing the potential to engage learners. Thus, when using these rubrics to choose appropriate OER, users need to consider these pedagogical guidelines.

Third, the findings could help users in selecting appropriate rubrics for OER evaluation tasks. In particular, users can choose rubrics whose content, development processes, and application contexts align most with the purpose of their evaluation. For example, users may consider whether they intend to evaluate resources for a particular discipline or for multiple disciplines, whether they wish to focus more on the reusability of resources or on the alignment of resources with educational standards, whether they need detailed scoring guides or more freedom to make choices, and whether they need training and/or support from rubric developers.

Fourth, this research could facilitate the future development of rubrics by clarifying what is common and what is lacking in existing rubrics. For example, rubric developers should consider testing and iteratively revising their rubrics as well as provide a rating scale, scoring guide, and other training material.

Finally, this research contributed to the improvement of OER quality. In particular, our research distilled common indicators for evaluating quality in OER, and helped illustrate what dimensions/aspects of quality have not been met by a certain resource, thereby providing suggestions on how to improve the OER.

Acknowledgements

References

Abramovich, S., Schunn, C. D., & Correnti, R. J. (2013). The role of evaluative metadata in an online teacher resource exchange. *Educational Technology Research and Development, 61*(6), 863-883.

Achieve, A. (2011). *Rubrics for evaluating open education resource (OER) objects.* Washington, D.C.: Achieve, Inc. Retrieved January 9, 2013, http://www.achieve.org/achieve-oer-rubrics

Achieve, A. (2014). *Educators evaluating quality instructional products.* Retrieved April 9, 2014, http://www.achieve.org/EQuIP

Arter, J., & Chappuis, J. (2006). *Creating & recognizing quality rubrics.* Pearson, Columbus.

Akpinar, Y. (2008). Validation of a learning object review instrument: Relationship between ratings of learning objects and actual learning outcomes. *Interdisciplinary Journal of Knowledge & Learning Objects*, 4, 291-302.

Atkins, D. E., Brown, J. S., & Hammond, A. L. (2007). *A review of the open educational resources (OER) movement: Achievements, challenges, and new opportunities.* Retrieved January 9,2013, http://www.hewlett.org/Programs/Education/OER/OpenContent/Hewlett+OER+Report.htm

Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly, 7*(1), 54-74.

Bethard, S., Wetzer, P., Butcher, K., Martin, J. H., & Sumner, T. (2009, June). Automatically characterizing resource quality for educational digital libraries. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (pp. 221-230). ACM.

Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., Lynch, C. A., Oblinger, D. G., Pea, R. D., Salen, K., Smith, M. S., Szalay, A. (2008). Fostering learning in the networked world: The cyberlearning opportunity and challenge, a 21st century agenda for the National Science Foundation. *Report of the NSF Task Force on Cyberlearning.* Virginia, US: NSF.

Bresciani, M. J., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K., & Hickmott, J. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation, 14*(12), 1-7.

Bundsgaard, J., & Hansen, T. I. (2011). Evaluation of learning materials: A holistic framework. *Journal of Learning Design, 4*(4).

Colton, D. A., Gao, X., Harris, D. J., Kolen, M. J., Martinovich-Barhite, D., Wang, T., & Welch, C. J. (1997). *Reliability issues with performance assessments: A collection of papers.* ACT Research Report Series, 97-3.

Custard, M., & Sumner, T. (2005).Using machine learning to support quality judgments. *D-Lib Magazine, 11* (10). Retrieved from http://www.dlib.org/dlib/october05/custard/10custard.html

DLESE (2004). *DLESE Reviewed Collection (DRC) best practices.* Retrieved 2014 from http://www.dlese.org/Metadata/collections/drc-best-practices.php

Fitzgerald, M. A., & Byers, A. (2002). A rubric for selecting inquiry-based activities. *Science Scope, 26*(1), 22-25.

Fitzgerald, M., Lovin, V., & Branch, R. M. (2003). The gateway to educational materials: An evaluation of an online resource for teachers and an exploration of user behavior. *Journal of Technology and Teacher Education, 11*(1), 21–51.

Haughey, M., & Muirhead, B. (2005).Evaluating learning objects for schools. *E-Journal of Instructional Sciences and Technology, 8*(1). Retrieved from http://ascilite.org.au/ajet/e-jist/docs/vol8_no1/fullpapers/eval_learnobjects_school.htm

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144.

Kastens, K. A., & Butler, J.C. (2001). How to identify the "best" resources for the reviewed collection of the Digital Library for Earth System Education. *Computers and the Geosciences, 27* (3), 375-378.

Kastens, K., Devaul, H., Ginger, K., Mogk, D., Defelice, B., DiLenonardo, C., Tahirkheli, S. (2005). Questions and challenges arising in building the collection of a digital library for education. *D-Lib Magazine, 11*(11). Retrieved from http://mirror.dlib.org/dlib/november05/kastens/11kastens.html

Kay, R., & Knaack, L. (2007). Evaluating the use of learning objects for secondary school science. *Journal of Computers in Mathematics and Science Teaching, 26*(4), 261-289.

Khanna, P., & Basak, P. C. (2013). An OER architecture framework: Needs and design. *International Review of Research in Open & Distance Learning, 14*(1).

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy, 65*(23).

Kurilovas, E., Bireniene, V., & Serikoviene, S. (2011). Methodology for evaluating quality and reusability of learning Objects. *Electronic Journal of e-Learning*, *9*(1).

Leacock, T. L., & Nesbit, J. C. (2007). A framework for evaluating the quality of multimedia learning resources. *Educational Technology & Society, 10*(2), 44-59.

Leary, H., Giersch, S., Walker, A., & Recker, M. (2009). Developing a review rubric for learning resources in digital libraries. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (pp. 421-422). ACM.

Leary, H., Giersch, S., Walker, A., & Recker, M. (2011). Developing and Using a Guide to Assess Learning Resource Quality in Educational Digital Libraries. In Kuo Hung Huang (Ed.), *Digital Libraries - Methods and Applications* (pp. 181-196). InTech (ISBN: 978-953-307-203-6, Available from:http://www.intechopen.com/articles/show/title/developing-and-using-a-guide-to-assess-learning-resource-quality-in-educational-digital-libraries)

Merlot (2014). *MERLOT peer review process.* Retrieved April 9, 2014, http://info.merlot.org/merlothelp/index.htm#about_merlot.htm

Moskal, Barbara M. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation, 7*(3). Retrieved May 6, 2014 from http://PAREonline.net/getvn.asp?v=7&n=3

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10), 71-81.

Nesbit, J., Belfer, K., & Leacock, T. (2007). Learning object review instrument (LORI), Version 1.5. *E-Learning Research and Assessment (eLera) and the Portal for Online Objects in Learning (POOL).*

Nesbit, J. C., & Li, J. (2004). Web-based tools for learning object evaluation. *Proceedings of the International Conference on Education and Information Systems: Technologies and Applications*, 2, 334-339.

Pérez-Mateo, M., Maina, M. F., Guitert, M., & Romero, M. (2011). Learner generated content: Quality criteria in online collaborative learning. *European Journal of Open, Distance and E-Learning.* http://www.eurodl.org/materials/special/2011/Perez-Mateo_et_al.htm

Porcello, D., & Hsi, S. (2013). Crowdsourcing and Curating Online Education Resources. *Science, 341*(6143), 240-241. http://www.sciencemag.org/content/341/6143/240.full

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing, 15*(1), 18-39.

Rodríguez, J. S., Dodero, J. M., & Alonso, S. S. (2011). Ascertaining the relevance of open educational resources by integrating various quality indicators. *RUSC: Revista de Universidad y Sociedad del Conocimiento, 8*(2).

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation, 11*(10).

U.S. Department of Education, Office of Educational Technology (2010). *Transforming American Education: Learning Powered by Technology*. Washington, D.C.

Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: Computer-mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications, 25*(3), 198-205.

Wetzler, P., Bethard, S., Leary, H., Butcher, K., Zhao, D., Martin, J. S., & Sumner, T. (2013).Characterizing and predicting the multi-faceted nature of quality in educational web resources. *Transactions on Interactive Intelligent Systems, 3*(3), 15:1-25.

Willner, P. (1997). Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. *Psychopharmacology, 134*(4), 319-329.

Wolfe, E. W., Kao, C., & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. *Written Communication, 15*(4).

Yuan, M., Recker, M. & Diekema, A. (2015). Do Rubrics Help? An Exploratory Study of Teachers' Perception and Use of Rubrics for Evaluating Open Educational Resource Quality. In *Proceedings of the 2015 Society for Information Technology & Teacher Education International Conference* (pp. 1314-1321). Las Vegas, NV: Association for the Advancement of Computing in Education (AACE).

Appendix

An Overview of Selected Rubrics

Rubric	Content			Rubric Development Processes	Application Context
	Quality Dimensions / Indicators	Rating Scale	Scoring Guide		
Rubric to evaluate OER objects (OER rubric; Achieve, 2011)	**8 quality indicators:** • Degree of alignment to standards • Quality of explanation of the subject matter • Utility of materials designed to support teaching • Quality of assessment • Quality of technological interactivity • Quality of instructional and practice exercise • Opportunities for deeper learning • Assurance of accessibility	0-3 & N/A (3: superior, 2: strong, 1: limited, 0: very weak/none, N/A: not applicable)	• Gives detailed scoring guide for all five scores	• Did not report empirical testing results • Revised several times	• Evaluate a variety of OER across different content areas and types • Apply to the smallest meaningful unit (e.g., a single lesson, a complete unit of study) • Generic rubric

Learning Object Evaluation Instrument (LOEI; Haughey & Muirhead, 2005)	**5 quality dimensions (15 quality indicators):** • Integrity • Usability • Learning • Design • Value	0-4 (0: absent, 1: weak, 2: moderate, 3: strong, 4: perfect)	• No scoring guide provided	• Reported empirical testing results	• Examine school-level content in K-12 contexts • More generic rubric
Learning Objects Quality Evaluation Model (Kurilovas, Bireniene, & Serikoviene, 2011)	**8 quality indicators:** • Technological criteria • Design and usability • Working stability • Architecture • Interactivity level • Language independence • Ease of use, intuitiveness • Open license, cost	Bad, poor, fair, good, excellent	• No scoring guide provided	• Reported empirical testing results	• Evaluate the quality of different learning objects • Generic rubric
Rubric to evaluate Learner Generated Content (LGC; Pérez-Mateo, Maina, Guitert & Romero, 2011)	**3 quality dimensions (42 quality indicators):** • Content • Format	Not specified	• No scoring guide provided	• Did not report empirical testing results	• Evaluate the resources generated by learners • Generic

	• Process				rubric
Learning Object Review Instrument (LORI, Leacock & Nesbit, 2007; Nesbit, Belfer, & Leacock, 2007; Vargo, Nesbit, Belfer, & Archambault, 2003)	**9 quality indicators:** • Content quality • Learning goal alignment • Feedback and adaptation • Motivation • Presentation design • Interaction usability • Accessibility • Reusability • Standards compliance	1-5 & N/A (1: poor, 5: excellent, N/A: rubric not applicable)	• Gives detailed scoring guide for point 1 and 5, and give example for point 3	• Reported empirical testing results • Revised several times	• Evaluate multimedia learning objects in an individual or collaborativ e way • Generic rubric
Rubric from Fitzgerald, Lovin, & Branch (2003)	**6 quality indicators:** • Accuracy • Appropriateness • Clarity • Completeness • Motivation • Organization	Not specified	• No scoring guide provided	• Did not report empirical testing results	• Review materials for inclusion in Gateway to Educationa l Materials website • Rubric specific to a website

Digital Library for Earth System Education (DLESE) review rubric (DLESE, 2004; Kastens & Butler, 2001; Kastens, 2005)	**7 quality indicators:** • Scientific accuracy • Importance or significance • Pedagogical effectiveness • Well-documented • Ease of use for students and faculty • Power to inspire or motivate students • Robustness as a digital resource	Not specified	• No scoring guide provided	• Did not report empirical testing results	• Review resources for inclusion in DLESE • Rubric specific to a website
Rubric from Merlot (www.merlot.org; Haughey & Muirhead, 2005)	**3 quality indicators:** • Quality of content • Potential effectiveness as a teaching-learning tool • Ease of use	1-5 (1: material not worthy of use, 5:excellent all around)	• No scoring guide provided	• Did not report empirical testing results	• Review learning objects for inclusion in website • Apply in both individual and collaborativ e way • Rubric specific to a website

| Collaborative Learning Object Exchange (CLOE; Haughey & Muirhead, 2005) | **17 quality indicators:**

 • The content of the learning object is accurate.

 • The use of technology is appropriate for this content.

 • The content is presented clearly and professionally (spelling/grammar, et cetera).

 • Appropriate academic references are provided.

 • Credits to creators are provided.

 • There are clear learning objectives.

 • The learning object meets the stated learning objectives.

 • The target learners are clearly identified (academic level addressed/technical ability/demographics).

 • There are clear instructions for using the learning | Not specified | • No scoring guide provided | • Did not report empirical testing results | • Peer review of materials for inclusion in the website

 • Include two review panels: instructional designers, subject matter experts

 • Rubric specific to a website |

object.

- The technology helps
  learners to engage
  effectively with the
  concept/skill/idea.

- The learning object
  provides an
  opportunity for
  learners to obtain
  feedback within or
  outside the learning
  object.

- The author provides
  evidence that the
  learning object
  enhances student
  learning.

- Pre-requisite
  knowledge/skills, if
  needed, are
  identified.

- The learning object
  stands alone and
  could be used in
  other learning
  environments.

- The learning object
  is easy to use (i.e.
  navigation, user
  control).

- The author indicates
  whether the learning
  object is accessible
  for learners with

	diverse needs. • Technical requirements for the learning object are provided. (Draft Review Guidelines 2003)				
Rubric to evaluate learning resources in digital libraries (Leary, Giersch, Walker, & Recker, , 2009; Leary, Giersch, Walker, & Recker, 2011)	**6 quality indicators:** • Content accuracy • Text clarity • Links in project • Project completeness, includes the state standard learning goal, assessment, etc. • Purpose of the project • Overall rating of the project	1-5 (1:low quality, 5:high quality)	• No scoring guide provided	• Reported empirical testing results	• Rubric specific to the Instruction al Architect website
Rubric to evaluate instructional materials (Fitzgerald & Byers, 2002)	**4 quality dimensions (23 quality indicators):** • Increase students' understanding of the science subject matter investigated • Gain an understanding of how scientists study	Not specified	• No scoring guide provided	• Did not report empirical testing results	• Review resources to support inquiry in the classroom • Rubric specific to science domain

	the natural world				
	• Develop the ability to conduct investigation				
	• Develop the habits of mind associated with science				
Educators Evaluating the Quality of Instructional Products (EQuIP; Achieve, 2014)	**4 quality dimensions for math and literacy (31 quality indicators for K-–2 literacy, 29 indicators for K3-– 5 literacy, and 19 indicators for math):** • Alignment to the depth of the Common Core State Standards (CCSS) • Key shifts in the CCSS • Instructional supports • Assessment **3 quality dimensions for science (18 quality indicators)** • Alignment to the Next Generation Science Standards	0-3	• Gives detailed scoring guide	• Did not report empirical testing results • Revised several times	• Evaluate instruction al materials for use in schools • Rubric specific to mathemati cs, literacy, and science domains

(NGSS)

- Instructional supports

- Monitoring Student Progress

Rubric from Custard & Sumner (2005)	**5 quality dimensions (16 quality indicators):** • Provenance • Description • Content • Social authority • Availability	Y/N and text entry	• No scoring guide provided	• Reported empirical testing results	• Evaluate resources in an automated way • Generic rubric
Rubric from Bethard, Wetzer, Butcher, Martin, & Sumner (2009); Wetzler et al. (2013)-	**7 quality indicators :** • Has prestigious sponsor • Content is appropriate for age range • Has sponsor • Identifies learning goals • Has instructions • Identifies age range • Organized for	Y/N	• No scoring guide provided	• Reported empirical testing results	• Evaluate resources in an automated way • Generic rubric

learning goals

Note. **If a rubric proposes both quality dimensions and quality indicators, then the quality dimensions are presented here. If a rubric only proposes quality indicators, then the quality indicators are presented. If a rubric was revised several times, then the quality indicators and rating scales were reported based on the latest version of the rubric.**

© Yuan & Recker

Athabasca University