Methods of Multiple Criteria Evaluation of the Quality of Learning Management Systems for Personalised Learners Needs

Eugenijus Kurilovas

Institute of Mathematics and Informatics, Vilnius, Lithuania
eugenijus.kurilovas@itc.smm.lt

Abstract. The main research object of the paper is investigation and proposal of the quality evaluation method suitable for the multiple criteria evaluation (decision making) and optimisation of learning software packages – Learning Management Systems (LMSs). Both LMSs general technological 'internal quality' and adaptation 'quality in use' evaluation criteria are analysed in the paper and incorporated into the comprehensive quality evaluation method. The LMSs quality evaluation criteria are further investigated as the possible LMSs optimisation parameters and the experts' additive utility function is explored to be applied to optimise LMSs according to personalised learners needs.

Keywords: multiple criteria evaluation, quality, learning management systems, adaptation, personalised learners needs, optimisation

1 Introduction: Problem of the Multiple Criteria Evaluation and Optimisation of Learning Software Packages

The problem of learning software packages' (such as LMSs) quality evaluation and optimisation is high on the European research and education agenda.

The paper is aimed to consider the problems of *expert* evaluation of *technological quality* of LMSs.

The basic *notions, principles and methods* applied in the paper are as follows.

Learning object (LO) is referred to as any digital resource that can be reused to support learning [13]. *LO repositories* (LORs) are considered here as properly constituted systems (i.e., organised LOs collections) consisting of LOs, their metadata and tools / services to manage them [8]. *LMSs* are considered here as specific information systems which provide the possibility to create and use different learning scenarios and methods [5]. *Quality evaluation* is defined here as the systematic examination of the extent to which an entity (part, product, service or organisation) is capable of meeting specified requirements [6].

Different scientific methods are used for quality evaluation of software. *Multiple criteria evaluation method* is referred to as the experts' additive utility function presented further in Section 5 including the alternatives' evaluation criteria, their ratings (values) and weights.

Expert evaluation is referred to as the *multiple criteria evaluation* of the learning software packages aimed at the selection of the best alternative based on score-ranking results. According to [1], if the set of decision alternatives is assumed to be predefined, fixed and finite, then the decision problem is to choose the optimal alternative or, maybe, to rank them. But usually the experts (decision

makers) have to deal with the problem of optimal decision in the multiple criteria situation where the objectives are often conflicting. In this case, according to [1], an optimal decision is the one that maximises the decision maker's utility.

The author applies here the software engineering principle (*Principle*) which claims that one should evaluate the software using the two different groups / types of evaluation criteria – 'internal quality' and 'quality in use' criteria. According to [2], '*internal quality'* is a descriptive characteristic that describes the quality of software independently from any particular context of its use, and *'quality in use'* is evaluative characteristic of software obtained by making a judgment based on criteria that determine the worthiness of software for a particular project or user / group. It is impossible to evaluate quality in use without knowing characteristics of internal quality [2].

The rest of the paper is organised as follows. Section 2 presents General 'internal quality' LMSs technological quality evaluation criteria, Section 3 – Adaptation 'quality in use' LMSs technological quality evaluation criteria, Section 4 – comprehensive technological quality evaluation tool (set of criteria) for LMSs, Section 5 – multiple criteria evaluation and optimisation of LMSs for the personalised learner needs. Conclusion and results are provided in Section 6.

2 General Criteria for Quality Evaluation of Learning Management Systems

Methodology of Technical Evaluation of LMSs is a part of the Evaluation of Learning Management Software activity undertaken as part of the New Zealand Open Source LMSs project [12]. The evaluation criteria in this methodology expand on a subset of the criteria, focusing on the technical aspects of LMSs [12]: (1) Overall architecture and implementation: Scalability of the system; System modularity and extensibility; Possibility of multiple installations on a single platform; Reasonable performance optimisations; Look and feel is configurable; Security; Modular authentication; Robustness and stability; Installation, dependencies and portability; (2) Interoperability: Integration is straightforward; LMS standards support (IMS Content Packaging, SCORM). The author has proposed to analyse also IMS Learning Design (LD) (together with LD compliant tools, e.g., RELOAD, LAMS v.2.0.3 together with Moodle v.1.8, MOT+, etc.) to create and reuse Units of Learning as well as IMS Common Cartridge [7] [9]; (3) Cost of ownership; (4) Strength of the development community (for open source products): Installed base and longevity; Documentation; End-user community; Developer community; Open development process; Commercial support community; (5) Licensing; (6) Internationalisation and localisation: Localisable user interface; Localisation to relevant languages; Unicode text editing and storage; Time zones and date localisation; Alternative language support; (7) Accessibility: Text-only navigation support; Scalable fonts and graphics; (8) Document transformation [9].

3 Adaptation Criteria for Quality Evaluation of Learning Management Systems

Graf and List paper [3] presents an evaluation of open source e-learning platforms / LMSs with the main focus is on adaptation issues – adaptability, personalisation, extensibility, and adaptivity capabilities of the platforms. Adaptation received very little coverage in e-learning platforms. An e-learning course should not be designed in a vacuum; rather, it should match students' needs and desires as closely as possible, and adapt during course progression. The extended platform will be utilised in an operational teaching environment. Therefore, the overall functionality of the platform is as important as the adaptation capabilities, and the evaluation treats both issues. There are only a few LMSs evaluations available in the current literature. Their main focus is on commercial products. In contrast, the work [3] is focused on open source products only. This research is focused on customisable adaptation only, which can be done without programming skills.

LMSs adaptation criteria are [3]: (1) Adaptability – includes all facilities to customise the platform / LMS for the educational institution needs (e.g., the language or the design); (2) Personalisation aspects – indicate the facilities of each individual user to customise his / her own view of the platform; (3) Extensibility – is, in principle, possible for all open source products. Nevertheless, there can be big differences. For example, a good programming style or the availability of a documented application programming interfaces are helpful; (4) Adaptivity – indicates all kinds of automatic adaptation to the individual user's needs (e.g., personal annotations of LOs or automatically adapted content).

The evaluation [3] is based on the qualitative weight and sum approach (QWS). QWS establishes and weights a list of criteria and is based on the use of symbols. There are six qualitative levels of importance for the weights, frequently symbols are used: (1) E = Essential; (2) * = Extremely valuable; (3) # = Very valuable; (4) + = Valuable; (5) | = Marginally valuable; and (6) 0 = Not valuable.

The weight of a criterion determines the range of values that can be used to measure a product's performance. For a criterion weighted #, for example, the product can only be judged #, +, |, or 0, but not *. This means that lower-weighted criteria cannot overpower higher-weighted criteria. To evaluate the results, the different symbols given to each product are counted. Example results can be 2*, 3#, 3| or 1*, 6#, 1+. The product can now be ranked according to these numbers. But the results are sometimes not clear. There is no doubt that 3*, 4#, 2| is better than 2*, 4#, 2| but it is not clear whether it is better than 2*, 6#, 1+. In the latter case further analysis has to be conducted.

In [3] the authors have adapted the QWS approach in a way where the essential criteria are assessed in a pre-evaluation phase. These minimum criteria cover three general usage requirements: an active community, a stable development status, and a good documentation of the platform. The fourth criterion incorporates the didactical objective and means that the platform's focus is on the presentation of content instead of communication functionalities.

At the beginning of the evaluation, the authors of [3] have chosen 36 platforms and evaluated these according to the minimum criteria have been selected in [3]. Nine platforms (ATutor 1.4.11, Dokeos 1.5.5, dotLRN 2.0.3, based on OpenACS 5.1.0, Ilias 3.2.4, LON–CAPA 1.1.3, Moodle 1.4.1, OpenUSS 1.4 extended with Freestyle Learning 3.2, Sakai 1.0, and Spaghettilearning 1.1) meet the criteria. Next, these nine platforms were tested in detail. A questionnaire and an example of a real life teaching situation, covering instructions for creating courses, managing users and simulating course activities, were designed and applied to each platform.

Finally, [3] established eight categories: communication tools, learning objects, management of user data, usability, adaptation, technical aspects, administration, and course management.

The evaluation results of the adaptation category are presented in Table 1.

Table 1. Platform / LMS Adaptation Evaluation Results [3]

	Adaptability	**Personalisation**	**Extensibility**	**Adaptivity**	**Ranking**
Max. values	*	#	*	*	
ATutor	|	#	#	|	3
Dokeos	|	0	*	+	2
dotLRN	+	+	*	0	2
Ilias	+	#	*	0	2
LON–CAPA	+	#	#	|	2
Moodle	#	+	*	|	1
OpenUSS	#	#	#	0	2
Sakai	0	0	*	0	3
Spaghettilearning	+	#	+	0	3

Examining the results from a vertical perspective, it can be seen that the adaptability and the personalisation subcategories yield a broad range of results. The majority of the platforms were estimated as very good with regard to extensibility. In contrast, adaptivity features are underdeveloped.

As a result, Moodle can be seen as the best LMS concerning adaptation issues. Moodle provides an adaptive feature called ‘lesson’ where learners can be routed automatically through pages depending on the answer to a question after each page. Furthermore, the extensibility is supported very well by a documented API, detailed guidelines, and templates for programming. Also adaptability and personalisation aspects are included in Moodle. Templates for themes are available and can be selected by the administrator. Students can choose out of more than 40 languages [3].

4 Comprehensive Technological Quality Evaluation Tool for Learning Management Systems

In their previous papers [7] [10] the author has provided conclusions on the analysis of the aforementioned LMSs quality evaluation methods [3] and [12]. While analysing these methods it has been necessary to exclude all evaluation criteria that do not deal directly with LMSs technological quality problems on the one hand, and to estimate interconnected / overlapping criteria on the other.

This analysis has shown that the both analysed LMSs technological evaluation methods have a number of limitations: (1) the method developed in [12] practically does not examine adaptation capabilities criteria, and (2) the method proposed by [3] insufficiently examines general technological criteria. Therefore, in the author's opinion, a more comprehensive tool / set of criteria for LMSs technological evaluation is needed. It should include general technological evaluation criteria based on modular approach and interoperability, as well as adaptation capabilities criteria. LMSs adaptation capabilities criteria should have the same weight as the other criteria [10]. According to the *Principle*, the comprehensive LMSs quality evaluation tool should include both 'internal quality' evaluation criteria and 'quality in use' evaluation criteria.

The comprehensive set of criteria / tool for LMSs technological evaluation is proposed in Table 2.

This tool provides the experts the clear instrumentality who (i.e., what kind of experts) should analyse what kind of LMSs quality criteria in order to select the best LMS software suitable for their needs.

Table 2. LMSs technological quality evaluation criteria [7] [10]

Internal quality (General) evaluation criteria	1. Overall architecture and implementation	Scalability
		Modularity (of the architecture)
		Possibility of multiple installations on a single platform
		Reasonable performance optimisations
		Look and feel is configurable
		Security
		Modular authentication
		Robustness and stability
		Installation, dependencies and portability
	2. Interoperability	Integration is straightforward
		LMS standard support
	3. Internationalisation and localisation	Localisable user interface
		Localisation to relevant languages
		Unicode text editing and storage
		Time zones and date localisation
		Alternative language support
	4. Accessibility	Text only navigation support
		Scalable fonts and graphics

Quality in use (Adaptation) evaluation criteria	5. Adaptability (facilities to customise for the educational institution's needs)	Language
		Design
	6. Personalisation aspects (facilities of each individual user to his/her own view of the platform)	
	7. Extensibility	Good programming style
		Availability of a documented API
	8. Adaptivity (all kinds of automatic adaptation to the individual user's needs)	Personal annotations of LOs
		Automatically adapted content

The main ideas for the constitution of this tool are to clearly divide LMSs quality evaluation criteria according to the scientific *Principle* as well as to ensure the comprehensiveness of the tool and to avoid the overlap of the criteria.

5 Multiple Criteria Evaluation and Optimisation of Learning Management Systems for the Personalised Learner Needs

5.1. Ratings of the Quality Evaluation Criteria

Scientists who have explored quality of software consider that there exists no simple way to evaluate functionality characteristics of internal quality of software. According to [2], it is a hard and complicated task, which requires relatively high time and labour overheads.

According to [14], each alternative in multi-criteria decision making problem can be described by a set of criteria. Criteria can be qualitative and quantitative. They usually have different units of measurement and different optimisation direction. The author proposes to use the *multiple criteria evaluation method* for the evaluation of learning software packages. This method is expressed here by the experts' additive utility function presented further in the Section and including the alternatives' *evaluation criteria, their ratings (values) and weights*.

The comprehensive tool / set of evaluation criteria suitable for the expert multiple criteria evaluation (decision making) of LMSs has been proposed earlier in Table 2. According to the multiple criteria evaluation method, we also need LMSs evaluation criteria ratings (values) and their weights.

The measurement criteria of the decision attributes' quality used in [3] and [12] are mainly qualitative and subjective. Decisions in this context are often expressed in natural language, and evaluators are unable to assign exact numerical values to the different criteria. Assessment can be often performed by linguistic variables, e.g., 'bad', 'poor', 'fair', 'good' and 'excellent' in [12] or 'not valuable', 'marginally valuable', 'valuable', 'very valuable', and 'extremely valuable / essential' in [3]. These values are imprecise and uncertain: they are commonly

called fuzzy values. Integrating these different judgments to obtain a final evaluation is not evident.

Therefore, [11] have proposed to use fuzzy group decision making theory to obtain final assessment measures.

First, linguistic variable values are mapped into triangular fuzzy numbers – TFNs (l, m, u) (see Table 3).

Table 3. Linguistic variables conversion into triangular fuzzy numbers

Linguistic variables	TFN
Excellent / extremely valuable, essential	(0.700, 0.850, 1.000)
Good / very valuable	(0.525, 0.675, 0.825)
Fair / valuable	(0.350, 0.500, 0.650)
Poor / marginally valuable	(0.175, 0.325, 0.475)
Bad / not valuable	(0.000, 0.150, 0.300)

After the defuzzification procedure which converts the global fuzzy evaluation results, expressed by a TFN (l, m, u), to a non-fuzzy value E, the following equation has been adopted by [11]:

$$E = [(u - l) + (m - l)] / 3 + l, \quad (1)$$

The non-fuzzy values E for all aforementioned linguistic variables calculated according to the equation (1) are presented in Table 4.

Table 4. Linguistic variables conversion into non-fuzzy values E

Linguistic variables	Non-fuzzy value E (rounded down)
Excellent / extremely valuable, essential	0.850
Good / very valuable	0.675
Fair / valuable	0.500
Poor / marginally valuable	0.325
Bad / not valuable	0.150

These non-fuzzy values are suitable to be applied to measure the ratings of the evaluation criteria of learning software packages such as LMSs and LORs.

The application of this method for evaluation of LORs quality has been presented by the author while implementing EdReNe [4] project. EdReNe brings together web-based repositories of LOs with content owners and other stakeholders within education in order to share, develop and document strategies, experiences, practices, solutions, advice, procedures, etc. on the organisation, structuring and functionality of repositories [4]. The LORs quality assurance strategies have been ranked the highest priority by the EdReNe experts.

5.2. Experimental Evaluation of Learning Management Systems

If we want to evaluate (or optimise) the technological quality of learning software package (e.g., LMS or LOR) for the particular learner needs (i.e., to personalise his / her learning process in the best way according to the prerequisites, preferred learning speed and methods, etc.), we should apply the experts' additive utility function using the different weights of evaluation criteria. The weight of the evaluation criterion reflects the experts' opinion on the criterion's importance level in comparison with the other criteria for the individual learner / user.

For the most simple (general) case, when all LMS evaluation criteria are of equal importance, the experts should consider the equal normalised weights $a_i = 0.125$ according to the normalisation requirement

$$\sum_{i=1}^{m} a_i = 1,\ a_i > 0, \tag{2}$$

A possible decision could be to transform multi-criteria task into one-criterion task obtained by adding all criteria together with their weights. It is valid from the point of view of the optimisation theory, and a special theorem exists for this case.

Therefore, here we can apply the experts' additive utility function

$$f(X) = \sum_{i=1}^{m} a_i f_i(X),\ \sum_{i=1}^{m} a_i = 1,\ a_i > 0. \tag{3}$$

The major is the meaning of the experts' additive utility function (3) the better LMS meets the particular learner needs.

The example of experimental evaluation of three popular open source LMSs ('ATutor', 'Ilias', and 'Moodle') according to the experts' additive utility function (3) is presented in Table 5.

The Table presents the values of the function (3), when the non-fuzzy values E for all linguistic variables in Table 4 are calculated according to the equation (1), and all LMS evaluation criteria are of equal importance $a_i = 0.125$ according to the normalisation requirement (2).

Table 5. LMSs technological evaluation summary (general case)

Technological evaluation criteria	ATutor	Ilias	Moodle
General 'internal quality' criteria ratings (values E)			
Architecture and implementation	0.500	0.325	0.850
Interoperability	0.675	0.675	0.500
Internationalisation and localisation	0.325	0.500	0.675
Accessibility	0.850	0.325	0.500
Interim evaluation rating	*2.350*	*1.825*	*2.525*

Adaptation 'quality in use' criteria ratings (values E)			
Adaptability	0.325	0.500	0.675
Personalisation	0.675	0.675	0.500
Extensibility	0.675	0.850	0.850
Adaptivity	0.325	0.150	0.325
Interim evaluation rating	*2.000*	*2.175*	*2.350*
Total evaluation rating	*4.350*	*4.000*	*4.875*
Value of f(X) (all weights = 0.125)	*0.5437*	*0.5000*	*0.6093*

These results mean that Moodle meets 60.93% quality in comparison with the ideal (less than 'good'), ATutor – 54.37% (more than 'fair') and Ilias – 50.00% ('fair'). According to this experimental evaluation results, Moodle is the best alternative (among the evaluated) from technological point of view in general case.

In more specific cases, e.g., if the experts would like to select the most suitable LMS for the students with special education needs / disabilities, they should choose higher weights for the particular criteria such as Accessibility and Personalisation (e.g., measuring weights $a_4 = 0.2$ and $a_6 = 0.2$). All the other criteria weights according to the normalisation formula (2) should be measured $a_i = 0.1$.

In this particular case the experts should find that, differently from the simple general case (see Table 5), both ATutor and Moodle seem to be the optimal alternatives for the learners with special education needs (see Table 6).

Table 6. LMSs technological evaluation summary for the learners with special education needs (using different weights)

Technological evaluation criteria	**ATutor**	**Ilias**	**Moodle**
General 'internal quality' criteria ratings			
Architecture and implementation $a_1 = 0.1$	0.0500	0.0325	0.0850
Interoperability $a_2 = 0.1$	0.0675	0.0675	0.0500
Internationalisation and localisation $a_3 = 0.1$	0.0325	0.0500	0.0675
Accessibility $a_4 = 0.2$	0.1700	0.0650	0.1000
Interim evaluation rating	*0.3200*	*0.2150*	*0.3025*
Adaptation 'quality in use' criteria ratings			
Adaptability $a_5 = 0.1$	0.0325	0.0500	0.0675
Personalisation $a_6 = 0.2$	0.1350	0.1350	0.1000
Extensibility $a_7 = 0.1$	0.0675	0.0850	0.0850
Adaptivity $a_8 = 0.1$	0.0325	0.0150	0.0325
Interim evaluation rating	*0.2675*	*0.2850*	*0.2850*
Total evaluation rating f(X)	*0.5875*	*0.5000*	*0.5875*

These results mean that LMS ATutor and Moodle meet 58.75% quality in comparison with the ideal for special needs students (something between 'fair' and 'good'), and Ilias – 50.00% ('fair').

6 Conclusion and Results

The proposed LMSs quality evaluation method represented by the experts' additive utility function is based one the transformation of the multiple criteria task into the one-criterion task obtained by adding all criteria values together with their weights. This method is suitable to apply for the practical expert evaluation of LMSs to meet the particular learner needs. Therefore, it is of practical importance for public and private sectors' experts (decision makers), software engineers, programmers and users.

Such approach has never been applied for solving the learning software packages evaluation and optimisation tasks before.

References

1. Dzemyda, G., Saltenis, V. Multiple Criteria Decision Support System: Methods, User's Interface and applications. Informatica. Vol. 5, No 1–2, pp 31–42 (1994)
2. Gasperovic, J., Caplinskas, A. Methodology to evaluate the functionality of specification languages. Informatica. Vol. 17, No 3, pp 325–346 (2006)
3. Graf, S., List, B. An Evaluation of Open Source E-Learning Platforms Stressing Adaptation Issues. Presented at ICALT (2005)
4. EdReNe. EU eContentplus programme's Educational Repositories Network project web site. http://edrene.org/ (2009)
5. Institute of Mathematics and Informatics. Research on digital learning tools and virtual learning environments implementation in vocational education. Scientific research report, p 80. http://www.emokykla.lt/lt.php/tyrimai/194 (2005)
6. ISO/IEC 14598-1:1999. Information Technology – Software Product Evaluation – Part 1: General Overview. First edition, 1999-04-15
7. Kurilovas, E. Interoperability, Standards and Metadata for e-Learning. In: G.A. Papadopoulos and C. Badica (Eds.): Intelligent Distributed Computing III, SCI 237, pp 121–130. Springer-Verlag Berlin Heidelberg (2009)
8. Kurilovas, E. Digital Library of Educational Resources and Services: Evaluation of Components". Informacijos mokslai (Information Sciences). Vilnius, 2007, Vol. 42–43, pp 69–77 (2007)
9. Kurilovas, E. Several aspects of technical and pedagogical evaluation of virtual learning environments. Informatics in Education, Vol. 4 (2), pp 215–252 (2005)
10. Kurilovas, E., Dagiene, V. Learning Objects and Virtual Learning Environments Technical Evaluation Criteria. Electronic Journal of e-Learning Vol. 7, Issue 2, 2009, pp 127–136. Available online at www.ejel.org (2009)
11. Ounaies, H.Z., Jamoussi, Y., Ben Ghezala, H.H. Evaluation framework based on fuzzy measured method in adaptive learning system. Themes in Science and Technology Education. Vol. 1, Nr. 1, 2009, pp 49–58 (2009)
12. Technical Evaluation of selected Learning Management Systems. https://eduforge.org/docman/view.php/7/18/LMS%20Technical%20Evaluation%20-%20May04.pdf (2004)
13. Wiley. D. A. Connecting Learning Objects to Instructional design Theory: a definition, a Metaphor, and a Taxonomy. Utah State University. http://www.reusability.org/read/ (2000)
14. Zavadskas, E. K., Turskis, Z. A New Logarithmic Normalization Method in Games Theory. Informatica. Vol. 19, No 2, pp 303–314 (2008)