DETERMINING EVALUATION CRITERIA AND SUB-CRITERIA FOR E-LEARNING SOFTWARE

Ahmad Fadli Saad[1], Rukaini Abdullah[2], Liyana Shuib[3]

[1,2,3]Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Email: afadlis@siswa.um.edu.my[1], rukaini@um.edu.my[2], liyanashuib@um.edu.my[3]

ABSTRACT

Today's demands for e-learning have led to the emergence of numerous and diverse e-Learning software (e-LS) products in the market. With such a myriad of choices, selecting an e-LS can be difficult. In any software evaluation process, evaluation criteria are important for correct selection to be made. However, in the case of e-LS selection, information about its evaluation criteria is lacking. Hence, a Delphi study was conducted to identify the evaluation criteria for e-LS. This paper presents the study and its results. Eleven criteria and 66 sub-criteria were identified from the literature. A questionnaire comprising the criteria as items was distributed to 31 experts in the first round. 16 sub-criteria were added by the experts. After two Delphi rounds, three criteria were considered as being extremely important and eight criteria as important. One sub-criteria was rejected as it did not achieve the majority of expert consensus. In total, 11 criteria and 81 sub-criteria were obtained from this study. The results of this study indicate that these criteria and sub-criteria are important in the evaluation of e-LS.

Keywords: e-Learning Software, Evaluation Criteria and Sub-criteria, e-Learning Software Evaluation, Commercial off the Shelf Software

1.0 INTRODUCTION

The explosive growth of e-Learning technology has motivated organizations to implement e-Learning within their organizations. e-Learning refers to technology that can support education and training [1]. It is commonly correlated with the field of Advanced Learning Technology (ALT), which deals with technologies and methodologies in learning that uses networking and/or multimedia technologies [2]. e-Learning has been adopted by participators as an alternative training method [3]. This technology has become a common means for delivering knowledge to learners who may come from many different organizations including academic institutions [4]. Since e-Learning can save money, employee transportation time, and other expenditure, it has become an alternative way of delivering on-the-job training for many organizations, where learners can access e-Learning material from their home or workplace [5]. e-Learning is commonly supported by e-Learning software (e-LS) that can be used to customize or develop e-Learning applications [6]. e-LS can be categorized into Learning Content Management Systems (LCMS), Learning Management Systems (LMS) and Content Management Systems [71]. Organizations may also choose software deployment tools, such as Java Enterprise Beans, Microsoft.NET, Borland Enterprise Server and IBM Optimiselt products [7]. With so many varieties of e-LS on the market, organizations have considerable options in selecting suitable e-LS. Moreover, the task of e-LS selection has become more complex due to the ongoing improvements in information technology [8]. Choosing suitable e-LS is crucial in the decision-making process since a new release of e-LS products may have criteria that are not available in the software that is currently being evaluated. The wrong choice of e-LS could ensue from a lack of knowledge of the criteria considered during the evaluation process. Hence, information about the evaluation criteria and sub-criteria must be updated and considered important in the evaluation and selection process of e-LS products.

The aim of this study is to identify a list of evaluation criteria and sub-criteria for the evaluation and selection of suitable e-LS. In order to obtain a list of e-LS evaluation criteria that are relevant and important, the views of experts will be very helpful. With this list, organizations are provided with adequate criteria that can assist them in the evaluation and selection process of e-LS.

2.0 RELATED WORKS

Many criteria can be used for software evaluation. These criteria are commonly found in software quality models (SQMs) and Commercial off the-Shelf (COTS) approaches. SQMs, including McCall, Boehm, Dromey and ISO9126-1, form the foundation of software product quality [20,21,71,72,73]. Product quality refers to a set of attributes of software product, where its quality is described and evaluated [21]. SQMs provide the criteria to evaluate the technical aspects of products, such as functionality, reliability, usability, efficiency, maintainability and portability. ISO9126-1 uses each technical aspect above as the criteria [21] to evaluate software quality. Definitions of each criterion in the ISO 9126-1 quality model as can be found in [21][36], which are:

i. Functionality: The capability of the software system to meet the stated function when the system is used under the specified conditions.

ii. Maintainability: The capability of the software products to be modified. Modifications may include corrections, improvements or adoption of the system to change in respect of the environment and in the requirements and functional specifications.

iii. Usability: The capability of the software system to be understood, learned and attractive to the user when used for the specified conditions.

iv. Reliability: The capability of the software system to maintain its level of performance under the stated conditions for a stated period of time.

v. Portability: The capability of the software products to be transferred from one environment to another.

vi. Efficiency: The capability of the software system to provide appropriate performance relative to the amount of resources used under the stated conditions.

Due to the generic characteristics, these criteria can be used to evaluate any software [23] including e-LS [29]. However, the criteria need to be modified to suit the specific features of e-LS [75].

e-LS is a type of COTS product [71]. COTS software is defined as commercial pieces of reused software that are developed and supported by outside vendors [74]. COTS software can be integrated into new systems or reused in other projects [74][84]. Unlike the criteria in SQMs, criteria, such as Cost, Vendor/Supplier, Risk and Uncertainty, Product Benefit and Organizational, are used in the evaluation of COTS products [16][17][36][43]. These criteria evaluate the non-technical aspects of software that sometimes deemed to be more important than the criteria in SQM. Cost is an important criterion to be included in the evaluation process because of the high investment cost to acquire software products by organizations [30][38][48]. In addition to Cost, Vendor criteria is also important as it evaluates the services provided by the vendor. Organizations have to consider whether there is a vendor lock on the particular software product provided, which may resulted in a substantial cost when switching vendor [72]. This would affect the future investment in a software, such as e-LS.

Besides cost and vendor, uncertainty and risk of acquiring software also need to be taken into account. Risk and uncertainty refer to the possibility that the software fails to meet its goal or causes loss when implemented [11]. In a software development project, the loss describes the impact on the project, which could be in the form of diminished quality of the end product, increased costs, delayed completion, or failure of the project [11]. Risks with system management exist when the operation is not under the organization's direct control as well as risk related to the vendors [40][79]. For some organizations, the uncertainty and risk criteria are among the important criteria in the evaluation and selection of software products [10, 11]. Both criteria are strictly connected, because an incomplete knowledge about the software product characteristics lead to a higher risk of making the wrong choice of software products [11, 12].

It is also important to ensure that the software products used can benefit the organization. Product Benefit, evaluates the benefits of a software product; for example, the availability of the software to increase user satisfaction and the availability of the software to facilitate ease of use in system development [50][51][66]. Lastly, any criteria related to the organization, such as organizational culture, resources, politics and user acceptance should also be included in the evaluation of software [30][45][61].

For COTS products, approaches like Social Technical Approaches to COTS Evaluation (STACE) include a portion of the SQM criteria and COTS criteria [37][44]. This shows that both criteria from the SQMs and COTS are important to evaluate any software either COTS or non COTS poducts. The extent to which these

criteria are relevant and important for e-LS evaluation is determined in a Delphi study involving e-LS experts, as presented in the next section.

3.0 METHODOLOGY

This section explains the methodology used in this study. The four research questions to be answered in this study are as follow:

(1) What criteria and sub-criteria from past studies are relevant for e-LS evaluation?
(2) What criteria and sub-criteria for e-LS evaluation might be added by experts ?
(3) What is the degree of consensus among the experts towards the criteria and sub-criteria?
(4) What criteria and sub-criteria are important in the evaluation of e-LS based on priority?

The Delphi method was used to identify and rank the criteria and sub-criteria, which was devised at the Rand Corporation in the 1950s [83]. This method is considered to be a reliable research method for problem solving, decision-making, and group consensus in a variety of areas including software selection [14-18]. It obtains, refines and communicates the informed judgement of experts [12]. It involves multiple iteration or rounds of questionnaire or other means of data collection with the researcher controlling the statistical group's responses and feedback [76]. Since the consensus about the relative importance of the criteria from respondents are needed in this study, the Delphi method was chosen [13]. It allows the participation of respondents comprising experts to obtain their consensus concerning the criteria for e-LS evaluation. It involves a round by round of Delphi survey. The processes in conducting the Delphi survey are depicted in Fig. 1.

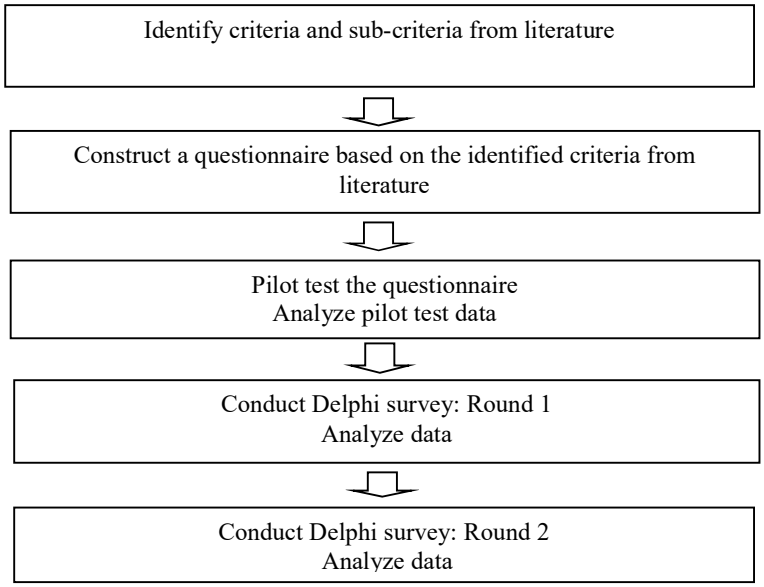

Fig. 1: Processes in conducting Delphi survey

3.1 Identify criteria and sub-criteria from literature

The search for e-LS evaluation criteria and sub-criteria was directed towards identifiying published papers in archival journals, conference proceedings and technical reports from electronic databases. The search was performed by browsing the Google Scholar, Web of Science and ISI web of Knowledge websites. Electronic databases, such as Elsevier's Science Direct, IEEE Xplore, ACM portal and Springer-Verlag's Link. Articles published in the proceedings of the IEEE on Software Engineering, Springer-Verlag, International conference on COTS-based software selection that were relevant to the area of this study were also included. e-LS criteria were collected from various sources – journals, reference books, websites and thesis dissertations. In addition, a manual reading of titles and abstracts of published papers, journals and conference papers that were potentially relevant was also obtained from the library to identify suitable source for literature. As the evaluation criteria

and sub-criteria were obtained from various sources, only papers that provide criteria and sub-criteria for software evaluation and e-Learning software were included to ensure the validity of the literature. A total number of more than 200 sources were reviewed. However only 50 of them are related and can be used to identify the selection criteria of e-LS. The criteria were listed and the references are cited. The criteria and sub-criteria obtained from the literature were compiled and used in the construction of the questionnaire for the Delphi study.

3.2 Respondents

The criteria and sub-criteria for e-LS were obtained from many sources of literature. As the criteria and sub-criteria were obtained from various resources, it must be validated by experts. This is to ensure that these criteria and sub-criteria are important for the evaluation of e-LS. An expert is any individual with relevant knowledge and experience of particular topic [80]. Respondents in this study were selected from organizations that implement e-Learning. They consist of decision makers, academicians and technical experts who have good knowledge and experience in e-Learning projects.

The minimum number of experts in the Delphi method is dependent on the study design. A Delphi panel that consists of a homogenous group of experts from the same general discipline area comprises ten to 15 participants [19,20]. In this study, there were ~~3~~ three homogenous groups comprising 10 decision-makers, 10 academicians and 11 technical experts. They were contacted via telephone or email. Of the initial 50 experts who agreed to participate, 31 responded to Round One and 31 to Round Two of the Delphi survey.

3.3 Constructing the questionnaire

The evaluation criteria and sub-criteria identified from the literature were compiled in the questionnaire to be given to the selected respondents. The questionnaire was divided into three sections. In Section A, the experts were asked about their demographic background such as type of organization, job function, educational attainment and years of experience in the field of Information Technology. Section B required the experts to rank the evaluation criteria that had been identified from the literature. The questions were recorded using a 5-point Likert-type scale that identified each criterion and sub-criterion of e-learning software products as "Extremely Important" (5), "Most Important" (4), "Moderately Important" (3),"Important" (2) and "Not Important" (1). Section C required the experts to provide new criteria and sub-criteria of e-LS products. The questionnaire with the initial list of criteria and sub-criteria from literature was sent to the experts in Round One of the survey. The information from the first round was collected and used to construct the questionnaire for Round Two of the survey. The questionnaire used in the second round is the same with the first round.

3.4 Conducting a Delphi survey

The Delphi method needs a minimum of two rounds of survey and most studies use only two or three rounds to obtain the consensus of experts [80]. In this present study, consensus was achieved after two rounds. In the first round, experts were asked to rank the criteria and sub-criteria identified from literature and provide additional criteria. The responses were analysed. In the second round, the same experts were involved in the first round. The goal of the second round was to reach a consensus concerning the relative importance of the criteria and sub-criteria that resulted from the first round. They were again asked to rank the criteria obtained from Round One and provide additional criteria and sub-criteria. Consensus was obtained in Round Two as no more criteria and sub-criteria were added by the experts in the second round of the survey.

3.5 Data analysis

Microsoft Excel 2010 and SPSS version 15.0 were used in the data analysis. Microsoft Excel 2010 was used to tabulate the criteria identified and collected from literature. Meanwhile, SPSS version 15.0 was used to analyse the consensus and priority of the criteria and sub-criteria among the experts. In order to analyze the data obtained from the Delphi survey, descriptive statistics were used. The analysis included:

i. Analysis of the consensus of experts concerning the criteria and sub-criteria.

The Inter Quartile Range (IQR) and Median were used to analyse the data obtained from the Delphi survey. The IQR was analysed to determine the level of the experts' consensus relevant to each criterion [73,76]. The calculation of the IQR revealed the relationships between each criterion and each expert which would lead to the interpretation that the experts have reached a consensus for each criterion [76]. On a five-point Likert scale, an IQR of 0 can be considered as high consensus while an IQR of 1 as good consensus, an IQR of 2 as a moderate consensus [70,77]. IQR of greater than 2.0 would indicate a disagreement between experts on their ratings [78]. Median score was checked to determine whether the criteria were accepted or rejected from the majority of experts [73,81,82]. On the 5-point Lickert scale also, the criteria that have a median value less than 3.5 were rejected as it does not achieve the consensus of the experts [81,82].

ii. Analysis of the priority of criteria according to experts.

To determine the priority of criteria for e-LS, the mean and standard deviation were analysed [13]. After analysis, the e-LS criteria were listed according to the sequence of importance obtained from Round Two. This led to the priority of criteria in view of the experts in this study.

4.0 RESULTS AND DISCUSSION

This section revisits the four research questions and discusses the results obtained in this study.

4.1 Evaluation criteria from literature.

From literature, 11 criteria and 66 sub-criteria were identified to be relevant for e-LS evaluation. C1 to C6 are criteria from SQMs. C7 to C11 are criteria commonly from COTS approaches. Table 1 lists the criteria obtained from the literature cited by 50 selected sources. The details of the criteria and sub-criteria are provided in Appendix I.

Table 1: List of criteria cited by 50 selected sources

Criteria	Source	Citation Number	% (n=50)
C1- Functionality	16,17,18,20,21,22,23,24,25,26,27,29,30,31,32, 33,34,35,36,37,38,39,40,41,53,54,55,56,58, 59,62,63,65	33	66%
C2- Usability	16,18,20,21,22,23,24,25,27,28,29,30,31,32,33, 34,35,36,37,38,38,39,40,41,52,53,55,56,57, 61,65	31	62%
C3-Portability	16,17,18,20,21,22,23,25,26,27,28,29,30,31,32, 33,34,35, 36,37,38,39,40,41,42,43,54,57,61	29	58%
C4-Maintainability	16,18,20,21,22,23,24,25,29,30,31,32,33,36,37, 38,39,41,57,58,62,63	22	44%
C5-Reliability	16,18,20,21,22,23,24,25,26,27,29,30,31,32,33, 36,37, 38,39,40,41	21	42%
C6-Efficiency	16,18,20,21,25,26,27,30,31,32,33,36,37,38,39, 41	16	32%
C7-Vendor	16,17,30,34,35,36,37,38,42,43,44,45,47,49,58	15	30%
C8-Cost	16,17,30,36,38,40,42,45,47,48,49,50,62,63	14	28%
C9-Risk &Uncertainty	17,19,40,48,51,63	6	12%
C10-Product Benefit	36,42,49,50,56,65	6	12%
C11-Organizational	30,37,45,60,64	5	10%

From Table 1, it can be seen that some criteria have more citation than others. For instance, the three most cited criteria are, Functionality, Usability, and Portability were found in more than half of the extracted papers. Meanwhile, Risk & Uncertainty, Product benefit and Organizational were only mentioned in a few papers. The most cited SQM criteria was Functionality and the least was Efficiency. The most cited COTS criteria was Vendor and the least was Organizational. This result answers the first research question on what criteria and

sub-criteria from past studies are relevant for e-LS evaluation. All these eleven criteria are relevant for e-LS evaluation.

4.2 Evaluation criteria and sub-criteria added by experts

As mentioned in section 3.4, experts were asked to provide additional criteria and sub-criteria in the Delphi survey. The survey results revealed that the experts did not add any new criteria. But they have added 16 new sub-criteria in Round One of the survey. The added sub-criteria are shown in Table 2.

Table 2: New sub-criteria added by experts

Criteria	**New Sub-Criteria**
Functionality	User/learner administration
Maintainability	Expansion Fault Software Error Preventing
Usability	Accessibility control
Reliability	Error Reporting
Portability	Standardability
Efficiency	Memory capacity
Cost	Marginal Cost
Vendor	-
Organizational	-
Product benefit	User Productivity Cost Saving After Sales Service
Risk Uncertainty	Frequency of software release Software Bug Unexpected cost Educational System Changed

Eight evaluation criteria received new sub-criteria, which are, Functionality, Maintainability, Usability, Reliability, Portability, Cost, Product Benefit and Risk and Uncertainty. For instance, experts added *User/learner administration* as a sub-criterion under Functionality. Two evaluation criteria - Vendor and Organizational had not received additional sub-criteria from the experts. In total, 66 sub-criteria were identified from literature and 16 sub-criteria were added by the experts. This result answers the second research question on criteria and sub-criteria that might be added by experts.

The following section presents the results of the experts' consensus on the criteria and sub-criteria.

4.2.1 Consensus of criteria and sub-criteria among the experts

Experts provided their opinions concerning the criteria and sub-criteria obtained in this study. The IQR and Median scores were analysed. Tables 3 to 13 present the consensus among the experts for all sub-criteria under each of the 11 criteria and the movement towards consensus in Round One and Round Two.

Table 3 shows the experts' consensus with respect to the ten sub-criteria of Functionality. The IQR and Median scores were obtained for each sub-criterion. After Round Two, the additional sub-criterion– *User/learner Administration* obtained High Consensus with IQR value of 0. From the remaining nine sub-criteria, eight sub-criteria achieved Good Consensus with IQR value of 1 while only one sub-criteria, *Personalization*, received Moderate Consensus with IQR value of 2.

There was also an improvement in consensus for the *SCORM Compliance* and *Accuracy* sub-criteria from Moderate Consensus (IQR value of 2) in Round One to Good Consensus (IQR value of 1) in Round Two. The other sub-criteria obtained consistent consensus with the same IQR in both rounds. The Median score for each sub-criterion was either 4 or 5. Thus all ten sub-criteria under Functionality were accepted due to the Median value being more than or equal to 3.5 as explained in [81][82].

Table 3: The consensus obtained with respect to the sub-criteria of Functionality

Sub Criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Suitability	1	4	Good Consensus	1	5	Good Consensus
Accuracy	2	4	Moderate Consensus	1	5	Good Consensus
Flexibility	1	4	Good Consensus	1	4	Good Consensus
Security	1	5	Good Consensus	1	5	Good Consensus
Interoperability	1	4	Good Consensus	1	4	Good Consensus
Pedagogical	1	4	Good Consensus	1	4	Good Consensus
Personalization	2	4	Moderate Consensus	2	4	Moderate Consensus
Learning Community	1	4	Good Consensus	1	5	Good Consensus
SCORM Compliance	2	4	Moderate Consensus	1	4	Good Consensus
UserLearner Administration (New Sub Criteria)	-	-	-	*0*	*4*	*High Consensus*

Table 4 displays the experts' consensus for nine sub-criteria under the Maintainability criteria. After Round Two, *Modularity*, and a newly added sub-criterion*, Expansion,* achieved High Consensus with IQR value of 0. The other two newly added sub-criteria, *Fault Software* and *Error Preventing* obtained Good Consensus with IQR value of 1. Other sub-criteria such as *Changeability, Stability, Analyzability* and *Scalability* also obtained Good Consensus with IQR value of 1. Only one criteria, *Testability*, received Moderate Consensus with IQR value of 2.

Table 4: The consensus obtained with respect to the sub-criteria of Maintainability

Sub Criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Changeability	1	4	Good Consensus	1	4	Good Consensus
Stability	2	4	Moderate Consensus	1	4	Good Consensus
Analyzability	2	4	Moderate Consensus	1	4	Good Consensus
Testability	2	4	Moderate Consensus	2	4	Moderate Consensus
Modularity	1	4	Good Consensus	0	4	High Consensus
Scalability	2	4	Moderate Consensus	1	4	Good Consensus
Expansion (New Sub Criteria)	-	-	-	*0*	*4*	*High Consensus*
Fault Software (New Sub Criteria)	-	-	-	*1*	*4*	*Good Consensus*
Error Preventing (New Sub Criteria)	-	-	-	*1*	*4*	*Good Consensus*

The level of consensus improved from Moderate to Good Consensus for the *Stability* and *Analyzability* sub-criteria and from Good to High Consensus for the *Modularity* sub-criterion after Round Two. The level of consensus for *Testability* was consistent with Moderate Consensus in both rounds.

The Median score for each sub-criteria under Maintainability was 4. Thus all nine sub-criteria for Maintainability were accepted because none of the Median was less than 3.5 [81,82].

Table 5 depicts the experts' consensus for the ten sub-criteria under Usability. After Round Two, one sub-criteria, *Learning Content*, and the newly added sub-criterion, *Accessibility Control*, received High Consensus with IQR value of 0. Seven sub-criteria – *Understandability, Learnability, Operability Customizability, Hypermediality, Support Tools* and *Presentation* obtained Good Consensus with IQR value of 1. One sub-criterion, *Learner Interface*, achieved Moderate Consensus with IQR value of 2.

There was an improvement in experts' consensus for Operability from Moderate (IQR value of 2) to Good Consensus (IQR value of 1). The level of consensus for the other eight sub-criteria- *Understandability, Learnability, Operability Customizability, Hypermediality, Support Tools, Presentation* and *Learning Content* were consistent with the same value of IQR in both rounds.

The Median value for *Learnability* also improved from 4 to 5.It shows that the number of experts who agreed increased. After Round Two, the Median results for each sub-criteria were either 4 or 5. Thus all ten sub-criteria for Usability were accepted as the Median value were not less than 3.5 [81,82].

Table 5: The consensus obtained with respect to the sub-criteria of Usability

Sub-criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Understandability	1	4	Good Consensus	1	4	Good Consensus
Learnability	1	4	Good Consensus	1	5	Good Consensus
Operability	2	4	Moderate Consensus	1	4	Good Consensus
Customizability	1	4	Good Consensus	1	4	Good Consensus
Hypermediality	1	4	Good Consensus	1	4	Good Consensus
Support Tools	1	4	Good Consensus	1	4	Good Consensus
Presentation	1	4	Good Consensus	1	4	Good Consensus
Learner Interface	2	4	Moderate Consensus	2	4	Moderate Consensus
Learning Content	0	4	High Consensus	0	4	High Consensus
Accessibility control (New sub-criteria)	-	-	-	*0*	*4*	*High Consensus*

The experts' consensus for each sub-criterion under the Reliability criteria is presented in Table 6. After Round Two, the newly added sub-criterion, *Error Reporting*, received High Consensus with IQR value of 0. Two sub-criteria – *Fault Tolerance* and *Backup Recovery* obtained Good Consensus with IQR value of 1. One sub-criterion – *Maturity* achieved Moderate Consensus with IQR value of 2. The consensus for this sub-criteria was consistent for both rounds.

The level of experts' consensus improved for *Fault Tolerance* and *Backup and Recovery* from Moderate Consensus (IQR = 2 in Round One) to Good Consensus (IQR = 1 in Round Two). For *Maturity*, the level of consensus was consistently moderate in both rounds.

The Median score for all sub-criteria was 4. Thus all four sub-criteria under Reliability were accepted as the Median value was not less than 3.5 [81,82].

Table 6: The consensus obtained with respect to the sub-criteria of Reliability

Sub Criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Maturity	2	4	Moderate Consensus	2	4	Moderate Consensus
Fault Tolerance	2	4	Moderate Consensus	1	4	Good Consensus
Backup Recovery	2	4	Moderate Consensus	1	4	Good Consensus
Error Reporting (New sub-criteria)	-	-	-	*0*	*4*	*High Consensus*

Table 7 shows the experts' consensus obtained for each sub-criterion under the Portability criteria. After Round Two, Conformance was the only sub-criterion that received High Consensus with IQR value of 0. The other sub-criteria including the added sub-criterion, *Standardability*, achieved Good Consensus

There was an improvement in the level of consensus for Conformance from Good Consensus (IQR =1 in Round One) to High Consensus (IQR = 0 in Round Two). The level of consensus was consistent for the other sub-criteria with no change of IQR value in both rounds.

There was also an improvement in the Median value from (Median Value = 3 in Round One) to (Median Value = 4 in Round Two) for Replaceability. This indicates the number of experts who agreed with Replaceability as an evaluation sub-criterion increased.

After Round Two, the Median result for each sub-criteria was 4. Thus all sub-criteria under Portability were accepted since none of the Median values was less than 3.5 [81, 82].

Table 7: The consensus obtained with respect to the sub-criteria of Portability

Sub Criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Adaptability	1	4	Good Consensus	1	4	Good Consensus
Installability	1	4	Good Consensus	1	4	Good Consensus
Replaceability	1	3	Good Consensus	1	4	Good Consensus
Conformance	1	4	Good Consensus	0	4	High Consensus
DBMS Standard	1	4	Good Consensus	1	4	Good Consensus
Middleware Standard	1	4	Good Consensus	1	4	Good Consensus
Standarability (New Sub Criteria)	-	-	-	*1*	*4*	*Good Consensus*

Table 8 displays the consensus obtained for the three sub-criteria under the Efficiency criteria. After Round Two, all three sub-criteria- Time Behaviour, Resource Behaviour and the added sub-criterion, *Memory capacity*, obtained High Consensus with IQR value of 0. The level of consensus for Time Behaviour improved from Good to High while the level of consensus for *Resource Behaviour* and *Memory Capacity* were consistent in both rounds. Since the Median score for all sub-criteria was more than 3.5, all three sub-criteria were accepted [81,82]

Table 8: The consensus obtained with respect to the sub-criteria of Efficiency

Sub Criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Time Behaviour	1	4	Good Consensus	0	4	High Consensus
Resource Behaviour	0	4	High Consensus	0	4	High Consensus
Memory Capacity *(New Sub Criteria)*	0	4	High Consensus	0	4	High Consensus

Table 9 shows the consensus obtained for each sub-criterion under the Cost criteria. After Round Two, five sub-criteria- *Licensing Cost, Implementation Cost, Maintenance Cost, Upgrade Cost, Training Cost* and the additional sub-criterion*, Marginal Cost*, achieved Good Consensus with IQR value of 1. Two sub-criteria – *Development Cost and Hardware Cost* received Moderate Consensus with IQR value of 2.

There was an improvement in the level of consensus for *Training Cost* from Moderate Consensus (IQR = 2 in Round one) to Good Consensus (IQR = 1 in Round Two). The other sub-criteria received consistent level of consensus where the same value of IQR was obtained in both rounds.

After Round Two, the Median score for each sub-criterion under Cost was 4. As the median scores were all greater than 3.5, those sub-criteria were accepted [81][82].

Table 9: The consensus obtained with respect to the sub-criteria of Cost

Sub Criteria	Round One			Round Two		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Licensing Cost	1	4	Good Consensus	1	4	Good Consensus
Development Cost	2	4	Moderate Consensus	2	4	Moderate Consensus
Implementation Cost	1	4	Good Consensus	1	4	Good Consensus
Maintenance Cost	1	4	Good Consensus	1	4	Good Consensus
Upgrade Cost	1	4	Good Consensus	1	4	Good Consensus
Hardware Cost	2	4	Moderate Consensus	2	4	Moderate Consensus
Training Cost	2	4	Moderate Consensus	1	4	Good Consensus
Marginal Cost (New Sub Criteria)	-	-		*1*	*4*	*Good Consensus*

The level of consensus obtained for each of the fourteen sub-criteria under the Vendor criteria are shown in Table 10. After Round Two, two sub-criteria – *Reputation and Training* received High Consensus with IQR value of 0. Eight sub-criteria–*Services, User Manual, Troubleshooting Guide, Maintenance and Upgrading, Communication, Demo, Technical and Business Skills and Past Business Experience* obtained Good Consensus with IQR value of 1. Four sub-criteria – *Support and Consultancy, Tutorial, Response Time and Length of Experience* obtained Moderate Consensus with IQR value of 2.There was no sub-criteria added by experts for this Vendor criteria. This indicates that experts agreed with the available sub-criteria from literature.

There was an improvement in the level of consensus for Reputation and Training which changed from Good Consensus (IQR = 1 in Round One) to High Consensus (IQR = 0 in Round Two) while for Services and User Manual, the change was from Moderate to Good Consensus. The level of consensus for the other sub-criteria were consistent where the IQR value remain unchanged in both rounds.

After Round Two, all fourteen sub-criteria under Vendor were accepted due to the Median scores being not less than 3.5 [81][82].

Table 10: The consensus obtained with respect to the sub-criteria of Vendor

Sub Criteria	Round 1			Round 2		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Reputation	1	4	Good Consensus	0	4	High Consensus
Support and Consultancy	2	4	Moderate Consensus	2	4	Moderate Consensus
Services	2	4	Moderate Consensus	1	4	Good Consensus
User Manual	2	4	Moderate Consensus	1	4	Good Consensus
Tutorial	2	4	Moderate Consensus	2	4	Moderate Consensus
Troubleshooting guide	1	4	Good Consensus	1	4	Good Consensus
Training	1	4	Good Consensus	0	4	High Consensus
Maintenance and Upgrading	1	4	Good Consensus	1	4	Good Consensus
Communication	1	4	Good Consensus	1	4	Moderate Consensus
Demo	1	4	Good Consensus	1	4	Good Consensus
Response time	2	4	Moderate Consensus	2	4	Moderate Consensus
Length of Experience	2	4	Moderate Consensus	2	4	Moderate Consensus
Technical and Business Skills	1	4	Good Consensus	1	4	Good Consensus
Past Business Experience	1	4	Good Consensus	1	4	Good Consensus

Table 11 shows the consensus obtained for each sub-criterion under the Organizational criteria. After Round Two, all sub-criteria obtained Good Consensus with IQR value of 1. There was no additional sub-criteria for Organizational.

The level of consensus was consistent for all sub-criteria as indicated by the same IQR value in both rounds. There was also an improvement in the Median value for *Organizational Resource* from (Median Value = 3 in Round One) to (Median Value = 4 in Round Two).This shows that the number of experts who agreed with *Organizational Resource* as a sub-criteria increased.

After Round Two, the Median result for each of the sub-criteria was 4. However, for *User Acceptance*, the Median result was less than 3.5. This criterion did not obtain the majority of expert consensus as the Median gained was only 3. This shows that experts in this study did not consider *User Acceptance* as an important sub criteria in the evaluation of e-LS. Therefore it was rejected from the e-LS sub-criteria list [81][82]. The other four sub-criteria were accepted.

Table 11: The consensus among expert obtained with respect to the sub-criteria of Organizational

Sub Criteria	Round 1			Round 2		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Organizational Culture	1	4	Good Consensus	1	4	Good Consensus
Organizational Resource	1	3	Good Consensus	1	4	Good Consensus
Organizational Change	1	4	Good Consensus	1	4	Good Consensus
Organizational Politics	1	4	Good Consensus	1	4	Good Consensus
User Acceptance	1	3	Good Consensus	1	3	Good Consensus

Next, Table 12 shows the consensus obtained for each sub-criterion under the Product Benefit criteria. After Round Two, *User Productivity*, an additional sub-criterion, obtained High Consensus with IQR value of 0.

Another two additional sub-criterion, *After Sales Service,* received Good Consensus while *Cost Saving* achieved Moderate Consensus. Two sub-criteria, User Satisfaction and Ease of Use both achieved Moderate Consensus.

After Round Two, the level of consensus for *User Satisfaction* and *Software Ease of Use* was consistent with unchanged IQR value in both rounds. With a Median value of 4 for each of the sub-criteria, all sub-criteria under Product Benefit were accepted as the scores were greater than 3.5 [81][82].

Table 12: The consensus among expert obtained with respect to the sub-criteria of Product Benefit

Sub Criteria	Round 1			Round 2		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
User Satisfaction	2	4	Moderate Consensus	2	4	Moderate Consensus
Software Ease of Use	2	4	Moderate Consensus	2	4	Moderate Consensus
User Productivity (New sub-criteria)	-	-	-	0	4	*High Consensus*
Cost Saving (New Sub-criteria)	-	-	-	2	4	*Moderate Consensus*
After Sales Service (New Sub Criteria)	-	-	-	1	4	*Good Consensus*

Finally, Table 13 shows the consensus for each sub-criterion under the Risk and Uncertainty criteria. After Round Two, one sub-criterion - *Vendor Risk* obtained Good Consensus with IQR value of 1. The four additional sub-criteria also received Good Consensus. *Product/Technological Risk* and *Software Bug* both achieved Moderate Consensus with IQR value of 2.

The level of consensus was consistent for *Vendor Risk, Product/Technological Risk* and *Software Bug* as indicated by the unchanged IQR value in both rounds. After Round Two, the Median result for each sub-criterion was 4. Thus all seven sub-criteria were accepted.

Table 13: The consensus among expert obtained with respect to the sub-criteria of Risk and Uncertainty

Sub Criteria	Round 1			Round 2		
	IQR	Median	Level of consensus	IQR	Median	Level of consensus
Vendor Risk	1	4	Good Consensus	1	4	Good Consensus
Product/Technological Risk	2	4	Moderate Consensus	2	4	Moderate Consensus
Software Bug	2	4	Moderate Consensus	2	4	Moderate Consensus
Frequency of Software Release (New Sub-Criteria)	-	-		1	4	*Good Consensus*
Virus and SPAM (New Sub Criteria)	-	-		*1*	*4*	*Good Consensus*
Unexpected Cost (New Sub Criteria)	-	-		*1*	*4*	*Good Consensus*
Educational System Changed (New Sub Criteria)	-	-		*1*	*4*	*Good Consensus*

In summary, the results of the IQR analysis revealed that 15.85% of the eighty-two sub-criteria received High Consensus; 64.63% obtained Good Consensus while 19.51% achieved Moderate Consensus. None of the sub-criteria received Without Consensus. The level of the consensus for the criteria and sub-criteria based on IQR is depicted in Fig. 2.

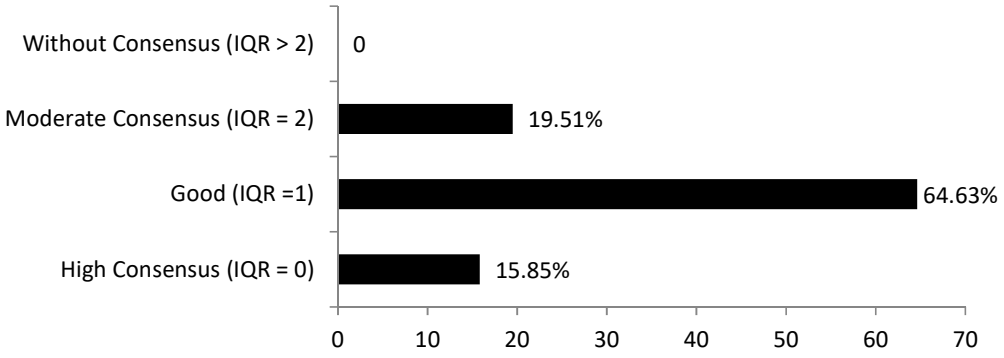

Fig. 2: Level of consensus for the sub-criteria based on IQR

All 16 new sub-criteria provided by the experts received between High and Moderate Consensus. Six sub-criteria such as *User/learner Administration*, *Expansion*, *User Productivity*, *Accessibility Control, Memory Capacity* and *Error Reporting* obtained High Consensus (IQR value of 0). The others obtained Good Consensus with IQR value of 1 except for *Cost Saving* which received Moderate Consensus. As mentioned in section 4.2.1, out of 82 sub-criteria, only one sub-criterion, *User Acceptance* was rejected as it did not achieve the majority of expert consensus. The remaining 81 sub-criteria achieved the majority of expert consensus and can be considered in the evaluation of e-LS. These results answer the third research question regarding the degree of consensus among the experts towards the criteria and sub-criteria.

4.3 The criteria and sub-criteria for e-LS based on priority

The mean and standard deviation were also analysed to identify the priority of the criteria and sub-criteria as ranked by experts. Of the eleven criteria identified in the literature, three criteria – Functionality, Usability and Reliability were ranked Extremely Important as shown in Table 14. The criterion with the highest mean average was Functionality with a mean average equal to 4.232. Next was Usability with a mean average of 4.087 and Reliability with a mean average equal to 4.056. The remaining eight criteria – Maintainability, Efficiency, Product Benefit, Cost, Vendor, Portability, Risk and Uncertainty and Organizational were ranked Important. The criterion with the lowest rank was Organizational with a mean average of 3.600.

Table 14: The ranking of criteria based on priority by experts

Criteria	Round 1 N=31		Round 2 N= 31			Scale
	Mean Average	Std Dev Average	Mean Average	Std Dev Average	Rank	
Functionality	4.104	0.777	4.232	0.701	1	Extremely Important
Usability	4.018	0.749	4.087	0.682	2	Extremely Important
Reliability	3.989	0.828	4.056	0.684	3	Extremely Important
Maintainability	3.989	0.801	3.996	0.668	4	Important
Efficiency	3.903	0.786	3.978	0.569	5	Important
Product Benefit	3.855	0.769	3.974	0.687	6	Important
Cost	3.959	0.739	3.968	0.715	7	Important
Vendor	3.809	0.798	3.929	0.752	8	Important
Portability	3.672	0.706	3.751	0.669	9	Important
Risk & Uncertainty	3.709	0.849	3.793	0.768	10	Important
Organizational	3.509	0.747	3.600	0.653	11	Important

This study revealed that all eleven criteria and sixty-six sub-criteria are important and can be used in the evaluation and selection of e-LS. This list provides a guideline to organizations on what criteria are important in the evaluation and selection e-LS. This result answers the fourth research question on what criteria and sub-criteria are important in the evaluation of e-LS based on priority. Appendix II shows the mean and standard deviation of each sub-criterion in detail.

5.0 CONCLUSION AND FUTURE WORK

We conducted a Delphi study involving experts, the first of its kind, to identify the evaluation criteria for e-LS. Experts' consensus on the criteria and sub-criteria were sought and then analysed based on the IQR and Median values. The experts' ranking of the criteria and sub-criteria based on priority were also analysed. This study identified a total of 11 criteria and 81 sub-criteria for e-LS evaluation. From the 81 sub-criteria, 16 were provided by the experts. Of the 11 criteria, three were ranked extremely important and eight were ranked important. The three criteria considered to be extremely important are Functionality, Usability and Reliability while the eight criteria that were ranked important are Maintainability, Efficiency, Product Benefit, Cost, Vendor, Portability, Risk & Uncertainty and Organizational. This list of criteria and sub-criteria obtained the consensus from the experts in this study and thus can be used in the evaluation and selection of e-LS. In the next study, we will investigate how these criteria can be integrated in a decision model to assist organizations in the evaluation and selection process of e-LS.

REFERENCES

[1] F. Dewever, "Opportunities for Open Source eLearning", *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, Vol. 1, No. 2, 2006, pp. 50-61.

[2] O. Floriana and Giovanni, "Discovering Student Models in e-learning System", *Journal of Universal Computer Science*, Vol. 1, No. 1, 2004, pp. 47-57.

[3] H. Fazlollahtabar, I. Mahdavi, "User/tutor Optimal Learning Path in e-learning Using Comprehensive Neuro-fuzzy Approach", *Educational Research Review*, Vol. 4, No. 2, 2009, pp. 142-155.

[4] L. Ismail, M. A. Serhani, S. Elnaffar, Y. Atif, "Semantic Grid-Based E-Learning Architecture", *Third International Conference on Semantics, Knowledge and Grid*, 2007, pp. 511-514.

[5] Y. Chen, C. Chang, "e-Homebook system: A Web-based Interactive Education Interface", *Computers and Education,* Vol. 11, 2005, pp. 197–211.

[6] D.Y. Shee, Y. S. Wang, "Multi-criteria Evaluation of the Web-based e-learning System: A Methodology Based on Learner Satisfaction and Its Applications", *Computers and Education*, Vol. 50, No. 3, 2008, pp. 894-905.

[7] H. Lee, "How Application Lifecycle Management Can Address eLearning Software Challenges", *International Journal of The Computer, the Internet and Management*, Vol. 12, No. 2, 2004, pp. 154 -161.

[8] H. Y. Lin, P. Y. Hsu and G. J. Sheen, "A Fuzzy-based Decision-making Procedure for Data Warehouse System Selection", *Expert Systems With Applications,* Vol. 32. No. 3, 2007, pp. 939-953.

[9] J. P. Carvallo, X. Franch and C. Quer, "Determining Criteria for Selecting Software Components: Lessons Learned", *IEEE Software,* Vol. 7, 2007, pp. 84-94.

[10] M. D. Tavana, D. Kennedy and P. Joglekar, "A Group Decision Support Framework for Consensus Rangking Technical Manager", *Omega,* Vol. 24, No. 5, 1996, pp. 523-538.

[11] E. Gülch, N. Al-Ghorani, B. Quedenfeldt and J. Braun, "Evaluation and Development of E-Learning Tools and Methods in Digital Photogrammetry and Remote Sensing for Non Experts from Academia and Industry", *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2012, Vol. 1, 1-6.

[12] E. R. Doke and N.E. Swanson, "Decision Variables for Selecting Prototyping in Information Systems Development: A Delphi Study of MIS Manager", *Information and Management*, Vol. 9, 1995, pp. 173-182.

[13] S. Dekleva and J. Zupancic, "Key Issues in Information Systems Management: a Delphi Study in Slovenia", *Information and Management*, Vol. 31, 1996, pp. 1-11.

[14] C. W. Holsapple and K. D. Joshi, "An Investigation of Factor That Influence the Management of Knowledge in Organization", *Journal of Strategic Information Systems*, Vol. 9, 2000, pp. 235-261.

[15] V. S. Lai, B. K. Wong and W. Cheung, "Group Decision Making in a Multiple Criteria Environment: A Case Using the AHP in Software Selection", *European Journal of Operational Research*, Vol. 137, No. 1, 2002, pp. 134-144.

[16] J. Kontio, C. Gianluigi and R. B. Victor, "Defining Factors, Goals and Criteria for Reusable Component Evaluation", in *Proceedings of the 1996 Conference of the Centre for Advanced Studies on Collaborative research*, IBM Press, 1996, p. 21.

[17] B. W. Boehm and D. Port, "Risk-Based Strategic Software Design: How Much COTS Evaluation is Enough?", in *Proceedings of the Third International Workshop on Economics-Driven Software Engineering Research. Toronto, Canada,* 2001.

[18] B. Kitchenham and S. L. Pfleeger, "Software Quality: The Elusive Target", *IEEE Software*, 1996, Vol. 1, 12-21.

[19] B. W. Boehm, "A Spiral Model of Software Development and Enhancement", *Computer*, 1988, Vol. 21, No. 5, pp. 61-72.

[20] R.G. Dromey, "Cornering the Chimera", *IEEE Software*, 1996, Vol. 1, pp. 33-43.

[21] ISO/IEC9126-1, "Software Engineering Product Quality-Part 1: Quality Model", International Standards Organisation, 2001.

[22] A. Mohamed, G. Ruhe and A. Eberlin,"COTS selection: past, present, and future", in *Engineering of Computer-Based Systems, 2007. ECBS'07. 14th Annual IEEE International Conference and Workshops on the,* IEEE, 2007, pp.103-114.

[23] X. Franch and J. P. Carvallo, "Using Quality Models in Software Package Selection", *IEEE Software,* Vol. 20, No. 1, 2003, pp. 34-41.

[24] S. Valenti, A. Cucchiarelli, M. Panti, "Computer based Assessment Systems Evaluation via the ISO9126 Quality Model", *Journal of Information Technology Education,* 2002.

[25] H. W. Jung, S. G. Kim and C. S. Chung, "Measuring Software Product Quality: A survey of ISO/IEC 9126", *IEEE Software*, Vol 5, 2004, 88-92.

[26] G. Grau, J. P. Carvallo, X. Franch, and C. Quer, "DesCOTS: a Software System for Selecting COTS Components", in *Euromicro Conference 2004, Proceedings 30th*, IEEE, 2004, pp. 118-126.

[27] T. Punter, R.V. Solingen and J. Trienekens, "Software Product Evaluation", *The Proceedings of the 4th IT Evaluation conference (EVIT-97)*, 30-31 Oct. 1997.

[28] M. F. Bertoa, J.M.Troya and A.Vallecillo, "Measuring the Usability of Software Components", *Journal of Systems and Software*, Vol. 79, No. 3, 2006, 427-439.

[29] B. B. Chua and L.E. Dyson, "Applying the ISO9126 Model to the Evaluation of an e-Learning System", *Proceedings of the 21st ASCILITE Conference*, 2004, pp. 5-8.

[30] J. P. Carvallo and X.Franch, "Extending the ISO/IEC 9126-1 Quality Model With Non-technical Factors for COTS Components Selection". In *Proceedings of the 2006 International Workshop on Software Quality* (2006, May), ACM, pp. 9-14.

[31] P. Robert, "Quality Requirements for Software Acquisition, "*Software Engineering Standards Symposium and Forum, 1997. Emerging International Standards. ISESS 9, Third IEEE International*, *IEEE*, 1997, pp.136-143.

[32] R. Pruengkarn, P. Praneetpolgrang, and A. Srivihok, "An Evaluation Model for e-Learning Websites in Thailand University", *Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on*. IEEE, 2005, pp. 161-162.

[33] B. Behkamal, M.Kahani, and M.K. Akbari, "Customizing ISO 9126 Quality Model for Evaluation of B2B Applications", *Information and Software Technology*, Vol. 51, No.3, 2009, pp. 599-609.

[34] S. Yacoub, A. Mili, C. Kaveri and M. Dehlin, "A Hierarchy of COTS Certification Criteria", in *Software Product Lines*, Springer US, 2000, pp. 397-412.

[35] M. Dehlin, S. Yacoub, A. Mili and C. Kaveri, "A Model for Certifying COTS Components for Product Lines", *in Workshop on Continuing Collaborations for Successful COTS Development, in Conjunction with the 22nd International Conference on Software Engineering (ICSE2000) Limerick, Ireland*, 2000.

[36] A. S. Jadhav, R. M. Sonar, "Evaluating and Selecting Software Packages: A review", *Information and Software Technology,* Vol. 51, No. 3, 2009, pp. 555-563.

[37] D. Kunda and L. Brooks, "Applying Social-Technical Approach for COTS Selection", in *Proceedings of the 4th UKAIS Conference*, 1999, pp. 552-565.

[38] A. S. Jadhav, R.M. Sonar, "Framework for Evaluation and Selection of the Software Packages: A Hybrid Knowledge based system approach", *Journal of Systems and Software*, Vol. 84, No. 8, 2011, 1394-1407.

[39] C. Quer, X. Franch, and X. Lopez-Pelegrin, "DesCOTS-SL: A Tool for the Selection of COTS Components", in *Requirements Engineering, 14th IEEE International Conference*, IEEE, 2006, pp. 365-366.

[40] M. Och, D. Pfahl, G. Chrobok-Diening and B.A. Nothhelfer-Kolb, "COTS Acquisition Process: Definition and Application Experience", *ISERN Report*, 2000.

[41] M. Ortega, M.Pér, and T. Rojas, "Construction of a Systemic Quality Model for Evaluating a Software Product", *Software Quality Journal*, Vol. 11, No.3, 2003, pp. 219-242.

[42] J. P. Carvallo, X. Franch and C. Quer, "Managing Non-technical Requirements in COTS Components Selection", *The Requirements Engineering, 14th IEEE International Conference*, 2006, pp. 323-326.

[43] M. S. Bandor, "Quantitative Methods for Software Selection and Evaluation", CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST, No. CMU/SEI-2006-TN-026, 2006.

[44] D. Kunda, L. Brooks, "Identifying and Classifying Processes (Traditional and Ssoft Factors) That Support COTS Component Selection: a Case Study", *European Journal of Information Systems*, Vol. 9, No. 4, 2000, pp. 226-234.

[45] B. W. Boehm, "A Spiral Model of Software Development and Enhancement", *Computer*, Vol. 21, No. 5, 1988, pp. 61-72.

[46] S. Sedigh-Ali, A. Ghafoor and R. Paul, "Software Engineering Metrics for COTS-based Systems", *Computer*, Vol. 34, No. 5, 2001, pp. 44-50.

[47] B. S. Sahay, A. K. Gupta, "Development of Software Selection Criteria for Supply Chain Solutions", *Industrial Management and Data Systems*, 2003, Vol. 103, No. 2, pp. 97-110.

[48] J. Du, J. Lu, D. Wu, H. Li and J. Li, " User acceptance of software as a service: Evidence from customers of China's leading e-commerce company, Alibaba", *Journal of Systems and Software*, 2013, Vol 86, No 8, pp. 2034-2044.

[49] C. A. Shoniregun and S.J. Gray, "Is E-learning Really the Future or a Risk?", *Ubiquity* 2003, April (2003), pp. 1-1.

[50] P. Y. Chau, "Factors Used in the Selection of Packaged Software in Small Businesses: View of Owner and Managers", *Information & Management*, Vol. 29, No. 2, 1995, pp. 71-78.

[51] K. A. Pituch and Y.K. Lee, "The Influence of System Characteristics on e-Learning Use", *Computers and Education*, Vol. 47, No. 2 , 2006, pp. 222-244.

[52] S. Patomviriyavong, B. Samphanwattanachai, T. Suwannoi, "eLearning Operational Risk Assesment and Management: A Case Study of the M.Sc in Management Program", *International Journal of Computers, the Internet and Management,* 2006.

[53] C. Ardito, M. F. Costabile, M. De Marsico, R. Lanzilotti, S. Levialdi, T. Roselli, V. Rossano, "An approach to Usability Evaluation of e-learning Applications", *Universal Access in the Information Society*, Vol. 4, No. 3, 2006, pp. 270-283.

[54] R. Lanzilotti, C. Ardito, M. F. Costabile and A. De Angeli, "eLSE methodology: a Systematic Approach to the e-learning Systems Evaluation", *Journal of Educational Technology and Society*, Vol.9, No. 4, 2005, pp. 42-53.

[55] S. Graf, B. List, "An Evaluation of Open Source E-Learning Platforms Stressing Adaptation Issues", *The Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, 2005.

[56] M. F. Costabile, T. Roselli, R. Lanzilotti, C. Ardito, V. Rossano, "A Holistic Approach to the Evaluation of e-learning Systems", in *Universal Access in Human-Computer Interaction. Applications and Services*, Springer Berlin Heidelberg, 2007, pp. 530-538.

[57] A. Koohang, "Expanding the concept of usability",*Informing Science*, *7*, 2004, pp. 129 - 141.

[58] E. W. Boot, J. J. Van Merriënboer, N.C. Theunissen, "Improving the Development of Instructional Software: Three Building-block Solutions to Interrelate Design and Production", *Computers in Human Behavior*, Vol. 24, No. 3, 2008, 1275-1292.

[59] K. M. Kapp, "Five Technological Considerations When Choosing an E-Learning Solution", *eLearn Magazine*, No. 6, 2003, pp. 1-5.

[60] K. K. Thyagharajan, and R.Nayak, "Adaptive Content Creation for Personalized e-Learning Using Web Services", *Journal of Applied Sciences Research*, Vol. 3, No.9, 2007, pp. 828-836.

[61] H. Mili, F. Mili, A. Mili, "Reusing Software: Issues and Research Directions", *Software Engineering IEEE Transactions on*, Vol. 21, No. 6, 1995, pp. 528-562.

[62] J. J. van Merriënboer, R. Martens, "Computer-based Tools for Instructional Design: An Introduction to the Special Issue", *Educational Technology Research and Development*, Vol. 50, No. 4, 2002, pp. 5-9.

[63] M. Kljun, J. Vicic, B. Kavsek, A. Kavcic, "Evaluating Comparisons and Evaluations of Learning Management Systems", in *Information Technology Interfaces, 2007, ITI 2007. 29th International Conference on*, IEEE, 2007, pp. 363-368.

[64] K. van den Berg, "Finding Open options. An Open Source software evaluation model with a case study on Course management Systems", Master's thesis, Tilburg University, *2005*.

[65] J. Du, J. Lu, D. Wu, H. Li and J. Li, "User acceptance of software as a service: Evidence from customers of China's leading e-commerce company, Alibaba", . *Journal of Systems and Software*, Vol. *86, No.*8, 2013, pp. 2034-2044.

[66] B. Mehlenbacher, L.Bennett, T.Bird, M. Ivey, J. Lucas, J. Morton, L. Whitman, "Usable e-learning: A Conceptual Model for Evaluation and Design", *Theories, Models, and Processes in HCI*, No. 4, 2005, pp. 1-10.

[67] J. McCall, P.K. Richards and G.F. Walters, "Factors in Software Quality: Final Report", Information Systems Programs, General Electric Company, 1977.

[68] B. W. Boehm, J. R. Brown and H. Kaspar, "Characteristics of Software Quality, Amsterdam: North-Holland, 1978.

[69] G.R. Dromey, "A Model for Software Product Quality", *Software Engineering, IEEE Transactions on*, Vol. 21, No. 2, 1995, pp. 146-162.

[70] H. A. Linstone, M. Turoff, "The Delphi Method: Techniques and Applications, Reading, MA:Addison-Wesley., 1975.

[71] K. Nagi, P. P. Suesawaluk and P.V. U-Lan, "Evaluating of eLearning Resource in A Learning Management System(LMS) - A Case Study of Moodle and Open Source Platform", *Special Issue of International Journal of the Computer, The Internet and Management*, 16**,** 2008, pp. 32.1-32.6.

[72] G. Siemens, "Learning or Management System? A Review of Learning Management System Review", University of Manitoba, 2006.

[73] A.R. Zakaria,H. Baharun and Z. Hasim, "Prospect of Physiotherapy Profession and Requirements for Changes in High School Curriculum", *The Role of Service in the Tourism & Hospitality Industry*, 77, 2015, pp. 77-83.

[74] F. Tawawneh, F. Baharom, J. Yahaya, F, Ahmad, "Evaluation and Selection COTS Software Process: The State of the Art", *International Journal on New Computer Architectures and Their Applications*, Vol. 1**,** 2011, pp. 344-357.

[75] Nagi, K. (2006). Solving ethical issues in e-learning. *Special Issue of the International Journal of the Computer, the Internet and Management*, *14*(7.17), 6.

[76] J. W. Murry and Hammons, J. O, "Delphi, a Versatile Methodology for Conducting Qualitative Research", *The Review of Higher Education,* Vol. 18, 1995, pp. 423-436.

[77] C. Vandelanotte, T. Dwyer, A. Van Itallie, C. Hanley, W. K. Murray, W.K, "The Development of an Internet Based Outpatient Cardiac Rehabilitation Intervention: A Delphi Study. BMC Cardiovascular Disorders", 2010, Vol.10, pp. 27.

[78] C. J. Kupec, "Training Needs for Project Management Professionals in Defense Industry", *Journal of Defense Studies and Resource Management,* Vol. 2, No. 2, 2013, pp. 2

[79] H. F. Lipson, N. R. Mead, A. P. Moore, "Can We Ever Build Survivable Systems from COTS Components?", in *Advanced Information Systems Engineering*, Springer Berlin Heidelberg, (2002, January), pp. 216-229.

[80] S. Thangaratinam and C.W, Redman, "The Delphi Technique", *The Obstetrician and Gynecologists*, Vol. 7, No. 2, 2005, pp. 120-125.

[81] C. G. Ononiwu, "A Delphi Examination of Inhibitors of The Effective Use of Process Industry Enterprise Resource Planning (Erp) Systems: A Case Study of New Zealand's Process Industry ", *The Electronic Journal Information Systems Evaluation*, Vol. 16, No. 2, 2013, pp. 114-131.

[82] M. Pichlak, "Innovation Generation Process and its Determinants", *International Journal of Contemporary Management*, Vol. 14, No. 1, 2015, pp. 51–66.

[83] C. C. Hsu and B. A. Sandford, "The Delphi Technique: Making Sense of Consensus", *Practical Assessment Research and Evaluation*, Vol. 12, No. 10, 2007, pp. 1-8.

[84] M.Morisio, C.B. Seaman,V.R.BasiliA.T. ParraS.E. Kraft and S.E. Condon," COTS-based software development: Processes and open issues", *Journal of Systems and Software*, Vol. 61, No. 3, 2002, pp. 189-199.

Appendix I

Relevant evaluation criteria for e-Learning software selection.

Criteria	Source
Functionality	16,17,18,20,21,22,23,24,25,26,27, 29,30,31,32,33,34,35,36,37,38,39,40,41, 53,54,55,56,58,59,62,63,65
Suitability	18,20,21,23,24,39,40,41
Accuracy	18,20,21,23,39,41
Flexibility	17,18,21,23,30,32,39,65
Security	16,17,18, 21, 23,24,25, 27, 30,31,32,33,34, 37, 39,41, 62
Interoperability	16, 18, 21,22,23,24,25, 27, 30, 32, 34,35, 37, 39,40,41
Pedagogical	52,53,55
Personalization	54, 59,62
Community	62,63
SCORM Compliance	58,62
Maintainability	16, 18, 20,21,22,23,24,25, 29,30,31,32,33, 36,37,38,39,41,57,58,62,63
Changeability	16, 18, 20,21,22,23, 25,26, 30,31,32,33,39,41
Stability	18,20,21, 23,25, 27,30,32,33,39,41,62
Analyzability	18,21,23, 25,30, 32, 33, 39, 41
Testability	16, 18, 20,21,22, 25, 27, 30, 32, 33,34, 39,41
Modularity	36,58,62,63
Scalability	36,38,57,62
Reliability	16, 18, 20,21,22,23,24, 25,26,27, 29,30,31,32,33, 36,37,38,39,40,41
Maturity	17,18, 21,22,23, 25,27,28,29, 32,33,39,41
Fault Tolerance	16,18,21,23,25, 27,30,32,33,39,41
Recoverability	16, 18,21,23,25, 27, 30,32,33,39,41
Usability	16,17,18,20,21,22,23,24,25,27,28,29,30,31,32,33,34,35,36,37,38,38, 39,40,41,52,53,55,56,57,61,65
Understandability	16,17,18,20,21,22,23,24,25,27,28,30,32,33,34,35,39,41,56
Learnability	16,18, 21, 23,24,25, 27,28,30,31,32,39,41,65
Operability	16,18,21,23,24,25, 27,28,30, 32,33,39,41
Customizability	18,21,23, 25,28,30, 32,33, 39,61,65
Hypermediality	52,53,55
Support Tools	52
Presentation	52,53,55
User Interface	36,38
Learning Contents	53,55,56
Portability	16,17,18,20,21,22,23,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40, 41,42,43, 54,57,61
Adaptability	16,18, 21,23,25,27,30, 32,33,34,35,36,37, 39,41,46,54,57,61
Installability	16,18,,21,23,25, 27,30,32,33,39,41
Conformance	16,18,21, 23, 25,30,32,39
Replaceability	18,21,23,25,27,30, 32,33,39,41
DBMS Standards	36,38
Middleware Standards	36,38
Efficiency	16,18,20,21,25,26,27,30,31,32,33,36,37,38,39,41
Time Behaviour	16,18,21,25,26,27,30,31,32,33,36,37,38,39,41
Recourse behaviour	16,18,21,25,26,27,30,31,32,33,36,37,38,39,41
Cost	16,17,30,36,38,40,42,45,47,48,49,50,62,63
Licensing Cost	30,37,42,47,50,62,63
Development Cost	30,42

Implementation Cost	30,42,45,47,48
Maintenance Cost	16,42,48
Upgrade Cost	47
Cost of Hardware	36,38
Training Cost	36,38
Vendor	16,17,30,34,35,36,37,38, 42,43,44,45,47,49,58
Vendor Reputation	30,34,35,42,44,45,49,58
Vendor Support &Training	17,34,37,42,43,44,45,47,49
Vendor Services	16,30,34,35,37,42,47,49
User Manual/Documentation	36,38
Tutorial	36,38
Troubleshooting Guide	36,38
Training	36,38
Maintenance and Upgrading	36,38
Communication	36,38
Demo	36,38
Response Time	36,38
Length of Experience	36,38
Technical and Business Skills	36,38
Past Business Experience	36,38
Organizational	30,37,45,60,64
Organizational Culture	45
Organizational Change	30,60
Organizational Politics	37,45
Organizational Resource	30,60
User Acceptance	64
Product Benefit	36,42,49,50,56,65
User Satisfaction	56,65
Software Ease of Use	49,50
Risk & Uncertainty	17,19,40,48,51,63
Vendor risk	17,40,48
Product/Technological Risk	17,19,40,48,51,63
Software Bugs and Errors.	19,51

Appendix II

Overall criteria and sub criteria ranked by respondents

		Round One		Round Two	
Citeria	Sub-criteria	Mean	Std Dev	Mean	Std Dev
Functionality	Suitability	4.194	0.703	4.387	0.715
	Accuracy	4.129	0.806	4.452	0.624
	Flexibility	3.742	0.773	3.742	0.855
	Security	4.419	0.719	4.548	0.675
	Interoperability	4.129	0.846	4.290	0.693
	Pedagogical	4.065	0.727	4.129	0.718
	Personalization	4	0.856	4.032	0.875
	Learning Community	4.194	0.792	4.484	0.569
	SCORM Compliance	4.065	0.772	4.161	0.688
New Sub criteria	User/Learner Administration	-		4.097	0.597
Maintainability	Changeability	4.097	0.597	4.194	0.703
	Stability	4.097	0.870	4.226	0.762
	Analyzability	3.903	0.789	4.129	0.718
	Testability	3.935	0.814	3.968	0.795
	Modularity	3.838	0.779	4	0.683
	Scalability	4	0.816	4.032	0.795
New Sub criteria	Expansion	-		3.839	0.454
New Sub criteria	Fault Software	-		3.774	0.560
New Sub criteria	Error Preventing	-		3.806	0.543
Usability	Understandability	4.226	0.805	4.387	0.667
	Learnability	4.290	0.824	4.484	0.626
	Operability	4.065	0.772	4.129	0.763
	Customizability	3.903	0.746	3.903	0.746
	Hypermediality	4.097	0.746	4.194	0.654
	Support Tools	3.871	0.763	3.871	0.718
	Presentation	3.871	0.670	3.935	0.727
	Learner Interface	3.903	0.789	4.065	0.854
	Leaning Content	3.935	0.629	3.968	0.547
New Sub criteria	Accessibility Control	-		3.935	0.512
Reliability	Maturity	3.903	0.944	4.065	0.814
	Fault Taularance	3.968	0.752	4.129	0.718
	Backup Recovery	4.097	0.789	4.194	0.749
New Sub criteria	Error Reporting	-		3.838	0.454
Portability	Adaptability	3.677	0.702	3.774	0.617
	Installability	3.806	0.792	3.903	0.831

	Replaceability	3.419	0.620	3.581	0.672
	Conformance	3.774	0.717	3.839	0.688
	DBMS Standard	3.709	0.692	3.806	0.703
	Middleware Standard	3.645	0.709	3.645	0.709
New Sub criteria	Standarability	-	-	3.709	0.461
Efficiency	Time Behaviour	3.839	0.820	4.06	0.629
	Resource Behaviour	3.968	0.752	4	0.516
New Sub criteria	Memory Capacity	-	-	3.871	0.562
Cost	Licensing Cost	3.903	0.700	3.935	0.679
	Development Cost	4.032	0.795	4	0.775
	Implement Cost	4.032	0.657	4.129	0.670
	Maintenance Cost	4.161	0.735	4.193	0.749
	Upgrade Cost	3.839	0.735	3.839	0.735
	Hardware Cost	3.774	0.762	3.968	0.752
	Training cost	3.968	0.795	3.935	0.727
New Sub criteria	Marginal Cost	-	-	3.742	0.631
Vendor	Reputation	3.839	0.583	4	0.577
	Support and Consultancy	3.871	0.885	3.935	0.892
	Services.	3.968	0.836	4.032	0.875
	User Manual	3.871	0.885	4.032	0.795
	Tutorial	3.806	0.873	4	0.775
	Troubleshooting Guide	3.774	0.762	3.903	0.746
	Training	3.871	0.670	4.032	0.605
	Maintenance and Upgrading	3.839	0.688	3.871	0.619
	Communication	3.839	0.779	4	0.730
	Demo	3.709	0.739	3.839	0.735
	Response Time	3.806	0.833	3.903	0.789
	Length of Experience	3.806	0.873	3.968	0.795
	Technical and Business Skills	3.742	0.815	3.774	0.805
	Past Business experience	3.581	0.958	3.709	0.783
Organizational	Organizational Culture	3.645	0.709	3.645	0.709
	Organizational Resource	3.194	0.601	3.581	0.620
	Organizational Change	3.613	0.667	3.709	0.588
	Organizational Politics	3.613	0.989	3.548	0.723
	User Acceptance	*3.484*	*0.769*	*3.516*	*0.626*
Product benefit	User Satisfaction	3.903	0.789	4	0.730
	Software Ease of Use	3.806	0.749	4	0.775
New Sub criteria	User Productivity	-	-	4	0.516
New Sub criteria	Cost Saving	-	-	4.065	0.814
New Sub criteria	After Sales Service	-	-	3.806	0.601
Risk Uncertainty	Vendor Risk	3.645	0.839	3.806	0.749
	Product/Technological Risk	3.742	0.815	3.903	0.789
	Software Bug	3.742	0.893	3.774	0.845

New Sub criteria	Frequency of Software Release	-	-	3.742	0.773
New Sub criteria	Software Bug	-	-	3.871	0.806
New Sub criteria	Unexpected Cost	-	-	3.709	0.783
New Sub criteria	Educational System Changed	-	-	3.742	0.631