Journal of Universal Computer Science, vol. 19, no. 7 (2013), 932-949
submitted: 31/8/12, accepted: 31/1/13, appeared: 1/4/13 © J.UCS

Assessment of Open-Ended Questions using a Multidimensional Approach for the Interaction and Collaboration of Learners in E-Learning Environments

Loc Phuoc Hoang
(Department of Computer Science, Faculty of Science, Khon Kaen University
Khon Kaen, Thailand
loc_hp@qtttc.edu.vn)

Ngamnij Arch-Int
(Department of Computer Science, Faculty of Science, Khon Kaen University
Khon Kaen, Thailand
ngamnij@kku.ac.th)

Abstract: Currently, the assessment of learners in conventional e-learning systems is only one dimension in which learners are required to produce answers, for example, by selecting multiple-choice, true/false, or matching answers or by giving short answers. This type of assessment still lacks interactions among the learners, and thus, it might not fully support learning. Many researchers have endeavored to propose an open-ended question method for evaluation, but their methods still focus on content assessment rather than learners' activities, which again lacks interactions among the learners. This paper concentrates on creating a new assessment method using open-ended questions with the aim of enhancing collaborations, activities and interactions of learners at the same time. The objectives are as follows: 1) to develop a process model for multidimensional assessment (M-DA) to enable effective learning; 2) to develop free-text answer assessments using a vector space model and a semantic extraction model; and 3) to develop an algorithm for evaluating learners based on a M-DA to encourage learners' activities. In addition, we created an environment for learners to be actively assessed and to interact with others when studying online. Two groups of parallel learners taking an e-course were tested on the two systems in a virtual learning environment. The results of the experiment noted that the system with multidimensional assessment showed a better outcome than the system without M-DA.

Keywords: E-learning collaboration, multidimensional assessment (M-DA), free-text answers assessment, vector space model, collaborative virtual environment.
Categories: L.0.0, L.0.1, L.2.0, L.3.5, L.3.6, L.6.2, I.2.7

1 Introduction

Currently, E-learning has successfully created new prospects for learners to study anywhere at any time [Islama, 11]. E-learning systems not only provide new possibilities for personalized learning at home or in the workplace but also reduce the requirements of costly traditional training for learners. However, barriers still exist in the assessment system; these barriers inhibit the efficiency of E-teaching and E-learning [Wong, 07; Assareh, 11].

The assessment in E-learning is facing a lack of quality assessment and interaction among learners, which hinders learners learning. The present E-learning assessment systems are generally in the form of multiple-choice, true/false, matching and short-answer items. To enhance the qualitative evaluation of the learners' knowledge and skills, some assessment techniques have been researched and published to support open-ended questions [Alfonseca, 04; Zhang, 08; He, 09; Hou, 10; Noorbehbahani, 11]. In this type of assessment, no answer choices are predetermined. Both the teacher and the learner are given an opportunity to build their own answer in the form of free text [Loc, 12]. Then, the system will check the learner's answer and score it by comparing it against the teacher's answer. In so doing, the teacher saves time in marking and scoring the learners' answers. However, these assessment methods focus only on the single-dimensional assessment of content rather than the learners' activities and interactions among the learners. Learners are required to answer according to what their teacher has taught, but they have no chance to analyze and comment on other learners' answers. This arrangement means that there is a lack of multidimensional assessment that correlates with learning in a present-day social network.

This research proposes a process model for multidimensional assessment (called M-DA) based on open-ended questions and free-text answers to enhance the study efficiency of learners in a virtual learning environment. We designed an assessment system in which teachers pose open-ended questions to which free-text answers can be given. Learners themselves are also encouraged to give free-text answers. The system then evaluates and scores each learner's answer by automatically comparing it with the teacher's answer. We relied on the vector space and semantic extraction model in this assessment. Additionally, the M-DA model is incorporated; this model is an active assessment method that evaluates learners based on their activities and knowledge comprehension. This assessment method aims to enhance the collaboration and interaction among learners that is necessary for both E-learning and social network learning systems. We propose a M-DA algorithm that, aside from allowing learners to answer the teacher's question, also enables learners to evaluate and comment on other learners' answers. If a learner can assess peers' answers and score them similarly to the system, he or she will receive an additional score. This procedure allows an interaction between learners and motivates learners to perform the test. Furthermore, seeing other learners' comments on one's answer means increased knowledge from the main part of the designed course content, and it motivates collaborative learning, which correlates to learning on a social network today.

The remainder of this paper is structured as follows: Section two describes related work regarding E-learning and assessment. Section three presents the assessment approach, including the presentation of a process model for multidimensional assessment, the conceptual framework of the M-DA system, an architecture for the assessment process, and a multidimensional assessment algorithm. The results of these experiments are described in section 4, and section 5 presents conclusions and a discussion of future work.

2 Related Work

The E-learning term has emerged for a long time, and it will eventually become an indispensable trend in modern education. E-learning is appealing to scholars' research in many countries worldwide. To satisfy learners' demands, E-learning has been developed using different technologies.

Using a social network is one of the developing trends behind current E-learning systems. Liccardi et al. [Liccardi, 07] indicated that a social network plays an important role in learners' knowledge acquisition. A social network is a good environment for learners' debates and discussion to discover knowledge about the content to be learned. Teachers can incorporate social networks into traditional class instruction. [Wang, 10] used software to analyze online courses in social networks to discover the position of learners in the virtual community. Their work also showed a relationship between the learners' position in the social network and the knowledge acquisition of the learners.

However, limitations of current social network learning exist regarding its capacity for online assessment or certifying learners' learning. There is a need to develop tools that support assessment and supervision in the virtual environment.

On-line learning collaboration designs have been researched and developed in recent years [Fardoun, 09; Hurtado, 11; Tissenbaum, 12; Caballé, 12]. These studies aim to enhance collaboration in a virtual learning environment. Additionally, methods and standards have been designed [Fardoun, 12; Alier, 12; Ozkan, 09] that can be used to build E-learning systems with flexibility and effectiveness in both technology and pedagogy.

The computer-assisted assessment of free-text answers has long been studied. Currently, most LMSs use simple question types, such as multiple choice, true/false, and matching. However, these types of questions are trivial assessments and are not accurate enough to measure the learners' knowledge. Many researchers focused on studying the automatic assessment of open-ended questions to enhance the quality of an assessment. In [Alfonseca, 04], the author focused on improving the basic BLEU algorithm by modifying the brevity penalty factor to solve the problem of learners' answers being written in short text; a word sense disambiguation (WSD) technique was also applied to enhance the assessment of the quality.

Zhang et al. [Zhang, 08] used the extracting multi-word method and a support vector machine to classify the documents. Another approach integrated latent semantic analysis (LSA) and n-gram co-occurrences to assess the learners' summary writings automatically. This approach assists the teacher in grading the learners' summaries effectively [He, 09]. Abdalgader et al. [Abdalgader, 10] proposed a short text similarity measure that integrated word sense disambiguation and synonym expansion to compute sentence similarities. A combination of the Part Of Speech tagging (POS) technique and support vector machines was used to assess the learners' answers. Notably, the precision rate was increased when Hou et al. associated this method with entropy to calculate the score of the learners' answers [Hou, 10]. Most of the assessment methods mentioned above focused only on enhancing the accuracy of an assessment on free-text answers. Learners' activities and interactions have not yet been studied and evaluated.

In [Noorbehbahani, 11], the modified BLEU algorithm (M-BLEU) was proposed to assess free text answers. The M-BLEU algorithm had four modifications: 1) M-BLEU used spell-check to check words when learners are typing and used synonymous expansion when matching n-grams. 2) The importance weight of every n-gram was recalculated to improve the precision of the M-BLEU algorithm. 3) Every learner's answer was compared with each reference answer (there were many reference answers designed for each question). A set of reference answers with a maximum score was then chosen (following the θ threshold) to calculate the similarity score of the question. 4) The authors calculated the maximum of a brevity penalty factor (BP) following a set of reference answers that obtained a maximum score, and they applied a maximum of the BP to select the best reference answer that was used to calculate the similarity of each learner's answer to the question.

The difference of syntaxes and size in comparison between the learners' answers and the teacher's answers is considerable; this difference directly influences the results of free-text answer assessments. Therefore, many authors proposed the BLEU algorithm and modified BLEU in [Alfonseca, 04; Noorbehbahani, 11]. The achieved results have a high correlation. However, teachers must create many different reference answers for any given question, and they will spend a substantial amount of time and effort in designing many reference answers to make it possible to choose the best answer that conforms to the syntaxes and the size of the learners' answers.

Castellanos-Nieves et al. [Castellanos-Nieves, 11] used semantic web technology to build both open and closed questions for assessment in E-learning systems. This method used a new technique to assess free-text answers. However, the authors did not compare the result of their method with previous methods, and it lacked multidimensional assessment.

Assessment is a fundamental task in an educational context; it is a pertinent phase that represents the quality of the output for an educational system. The automatic assessment of free-text answers in a virtual environment has two main goals. On the one hand, we would like to enhance the accuracy of assessment to support learners' grading, and it is worth noting that this area has attracted the attention of many researchers. On the other hand, the assessment aims to enhance active learning and comprehension. Learners' activities, interactions and collaborations should be enhanced in E-learning systems based on free-text answer assessments. This work remains as an open research area and requires the interest of more scholars. Based on this context, the paper focuses on creating a new assessment method to solve this problem.

3 The Assessment Approach

3.1 A Process Model for Multidimensional Assessment.

This section proposes a process model for M-DA that presents an outline of the assessment method to enhance effective learning. The process model is composed of several sub-processes, as shown in *Figure 1,* and it can be described as follows:

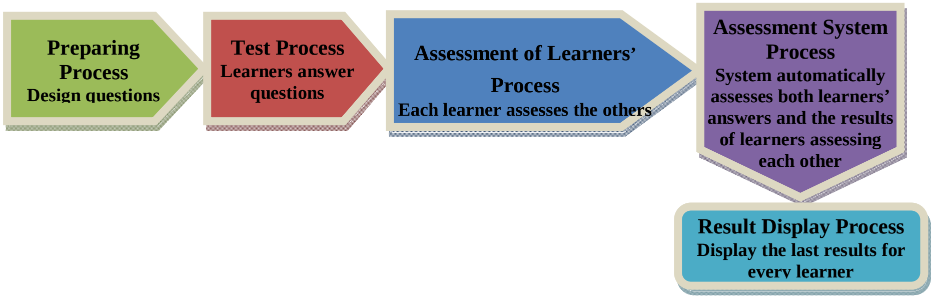

Figure 1: A Process Model for Multidimensional Assessment

1. **Preparation Process:** The teacher uses this process to provide questions, answers and time limits for learners' assessments and answers.
2. **Test Process:** Learners use this process to answer questions from the preparation process. In this process, learners must complete the questions within a limited amount of time.
3. **Assessment of a Learners' Process:** This process is used for each learner to assess other learners' answers during the time interval. This process aims to enhance interactions and collaborations between learners.
4. **Assessment System Process:** In this process, the system automatically assesses the learners' answers and the results of the learners assessing the others' answers. This process encourages learners to interact and study actively.
5. **Result Display Process:** For this final process, the system calculates and generates the last results and displays feedback for each learner.

The assessment of the learners' process and the assessment system process are described in detail in the next section.

3.2 The Abstract Conceptual Framework of the Multidimensional Assessment System

This paper proposes a M-DA method that not only assesses the learners' answers but also provides an environment that supports learners interacting by assessing other learners' answers.

The conceptual framework of M-DA has several components, as illustrated in *Figure 2*. Each component is described as follows:

3.2.1 Question or Topic Creation:

Teachers can use this function to create topics or open-ended questions that are used not only to assess learners when finishing the e-course but also to provide a topic or an exercise for learners to study and discuss with one another while they are learning.

3.2.2 The Answer Criteria Design:

This function is designed to enhance a method that creates many reference answers for each question, as mentioned previously in the related work section. The answer criteria are used as a guideline for learners to answer the question in a correct way and to avoid some trivial mistakes that can affect the assessment results by means of assessment techniques.

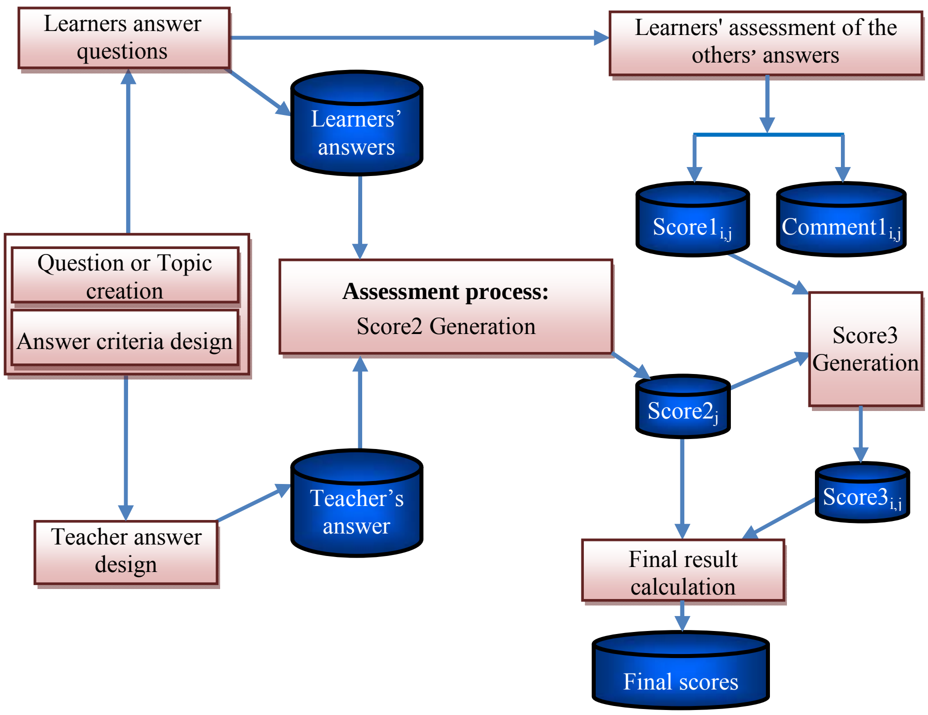

Figure 2: The abstract conceptual framework of the system

The answer criteria are designed to contain two parts:

- Approximate size: The estimated number of words for each answer.
- Description order of answer: The order of paragraphs, sentences or functions in each answer (instruction for each answer).

Example: The question: What is an operating system?
The answer criteria for this question contain:

- Approximate size: 80 words.
- Description of the answer: definition and functions of the operating system.

Using the answer criteria, learners must answer questions according to this answer criteria design. Hence, the learners' answers and the teacher's answer do not have much difference regarding the syntaxes and size. Therefore, the correlation of the learners' answers and teacher's answer are improved, and the scores of the assessment results are more accurate.

3.2.3 Teacher Answer Design:

This function allows teachers to write the answer to each question while conforming to the answer criteria design.

3.2.4 Learners Answer questions

This function allows learners to answer questions that conform to the answer criteria design.

3.2.5 Learners Assessment of the Others' Answers:

This function allows learners to assess and debate each other and give scores and comments on the others' answers. For example, $learner_i$ assesses $learner_j$'s answer with $Score1_{i,j}$ and $Comment1_{ij}$ (i, j=$\overline{1,n}, i \neq j$ because each learner cannot assess himself). $Comment1_{i,j}$ contains suggestions according to both deficient and correct information. The deficient information provides information that is lacking in the learner's answer, and the correct information provides information that is required in the learner's answer. We have $Score1_{j,i}$ and $Commnent1_{j,i,}$ respectively, when j=i and i=j. Through this work, learners can obtain knowledge in the e-course.

3.2.6 Assessment Process:

The system automatically assesses I answer and generates $Score2_j$ by matching $learner_j$*'s* answer with the teacher's answer. This function is presented in detail in Section 3.3.

3.2.7 Score3 Generation:

This function is used to generate $Score3_{i,j}$ by matching $Score1_{i,j}$ and $Score2_j$. The formulas for finding $Score3_{i,j}$ are given in formulas (5) and (6) in Section 3.5.

3.2.8 The Last Process Assessment Results:

This function is used to generate the last score for each $learner_i$ by calculating $Score2_i$ and $Score3_{i,j}$ with their respective coefficients. We can refer to formulas (7) and (8) in Section 3.5.

The overall algorithm for multidimensional assessment is illustrated in Section 3.4.

3.3 The Assessment Process on Free-text Answers

This section proposes the details of the assessment process on free-text answers as depicted in *Figure 3*. The design of the assessment process has several sub-processes, as follows:

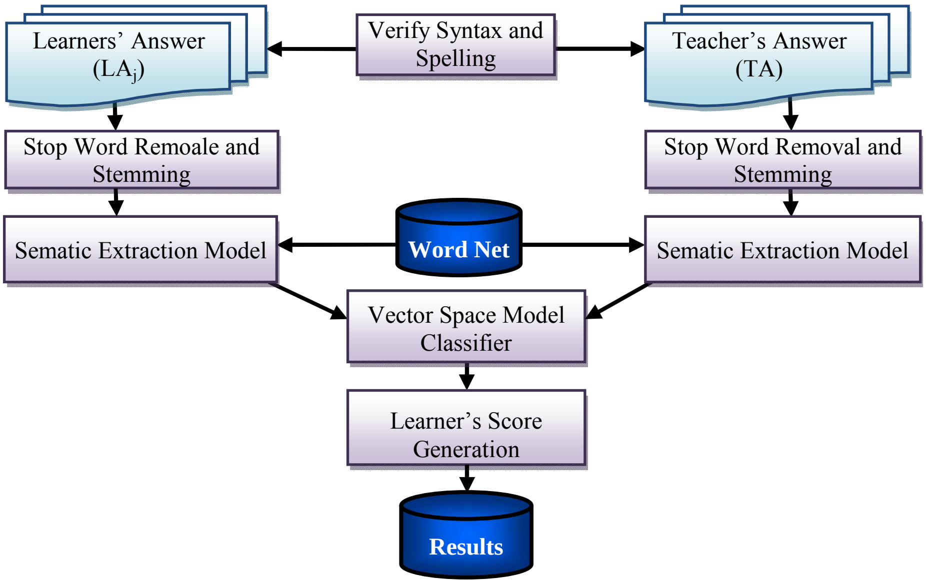

Figure 3: An overview framework for the automatic assessment process on free-text answers

3.3.1 Syntax and Spelling Verification:

This module is used for receiving and verifying the syntax and spelling of the students' answers and teacher's answer. This function employs the Java open source spell checker of Jazzy (http://jazzy.sourceforge.net/).

3.3.2 Learners' Answering and Teacher's Answering Processes:

These processes support learners in answering the question and support the teacher in designing the answer. Both the learners' and teacher's answers must conform to the answer criteria design, as defined in Section 3.2.2.

3.3.3 Stop Word Remove (SWR):

This module is used for removing the stop words, such as prepositions, conjunction, punctuations and special symbols, in the sentences of both the learner's and teacher's answers. These stop words must be removed before the free-text answers are processed in the next step.

3.3.4 Stemming:

This module is used for extracting the root words from words such as plurals and gerunds. Stemming is an important step in free-text assessment. This module can employ the Porter stemming algorithm[1].

3.3.5 Semantic Extraction Model:

This module is used to extract terms and term frequencies from a teacher's answer (TA) and a learner's answer (LA) and to find synonymous terms by employing the WordNet database[2]. This process is composed of several steps, as follows:

1. Transform terms from TA and each learner answer $LA_c \in LA_i$ into a matrix, and then count each term frequency that appears in TA and LA_c, respectively (see the example in *Table 1*).
2. Choose $term_i$ in the matrix that satisfies condition vectors $TA[term_i] \neq 0$ and $LA_c[term_i] = 0$ ($term_i$ appears in TA but not in LA_c).
3. Find synonymous terms for $term_i$ using WordNet; we now have a Synset that contains a set of synonymous terms for each $term_i$.
4. Compare each $term_j$ in the Synset with the $term_k$ that satisfies the conditions $LA_c[term_k] \neq 0$ and $TA[term_k] = 0$ ($term_k$ appears in LA_c but not in TA).
5. Choose the $term_j$ that is matched with $term_k$ such that $LA_c[term_k]$ = maximum and $TA[term_k] = 0$.
6. Assign $LA_c[term_i] = LA_c[term_k]$ and then remove $term_k$ from matrix. After this step, we obtain two expanded vectors, i.e., vectors LA_c and TA.

3.3.6 Vector Space Model Classifier and Learner's Score Generation:

This module is used to generate each learner's score by utilizing the vector space model formula[3] to calculate the similarity score between each learner's answer (LA_j) and each teacher's answer (TA), as shown in Formula (1).

$$Sim(LA_j, TA) = \frac{LA_j . TA}{\|LA_j\| \|TA\|} = \frac{\sum_k W_{j,k} . W_{TA,k}}{\sqrt{\sum_k W^2_{j,k}} \sqrt{\sum_k W^2_{TA,k}}} \qquad (1)$$

where each $W_{j,k}$ is a weight of the term T_k in LA_j and $W_{TA,k}$ is a weight of the term T_k in the teacher's answer. The weights $W_{j,k}$ and $W_{TA,k}$ are calculated using Formulas (2) and (3), respectively.

$$\mathbf{W_{j,k} = tf_{j,k} * idf_k \text{ or } W_{TA,k} = tf_{TA,k} * idf_k} \qquad (2)$$

---

[1] http://grepcode.com/snapshot/repo1.maven.org/maven2/gov.sandia.foundry/porter-stemmer

[2] http://wordnet.princeton.edu/

[3] http://en.wikipedia.org/wiki/Vector_space_model

where $tf_{j,k}$ is the frequency of the term T_k in the j-th answer and is calculated using Formula **(3)** and idf_k is an inverse document frequency of term T_k in the total number of answers that contain term T_k and is calculated using Formula (4).

$$\mathbf{tf_{j,k} = x_{jk}/N_j} \quad (3)$$

where x_{jk} is the frequency of the appearance of the term T_k in the j-th answer and N_j is the number of terms in the j-th answer.

$$\mathbf{idf_k = \log(N/n_k)} \text{ or } \mathbf{idf_k = \log(N/(n_k+1))} \quad (4)$$

where N is the total number of answers in the answers set and n_k is the number of answers in which term T_k appears.

Table 1 illustrates the steps to generate a score for $learner_c$ on a question. The last score of $learner_c$ calculated using Formula (1) is equal to 0.79371138. This value is converted into the marking scheme that has the maximum score = 10 $(0 \le score \le 10)$; therefore, the score of $learner_c$ = 7.94.

Unique Term	**Vectors**		**$TF_{c,k}$**		**IDF_k**	**Term weight – TF.IDF ($W_{c,k}$)**	
	TA	**LA_c**	**TA**	**LA_c**		**TA**	**LA_c**
E-learning	3	2	0.088	0.074	0.301	0.02649064	0.02227622
electronic	1	1	0.029	0.037	0.301	0.00872987	0.01113811
process	2	2	0.059	0.074	0.301	0.01776077	0.02227622
transfer	1	0	0.029	0	0.4771	0.013836516	0
skill	1	0	0.029	0	0.4771	0.013836516	0
Web-base	1	1	0.029	0.037	0.301	0.00872987	0.01113811
learning	2	2	0.059	0.074	0.301	0.01776077	0.02227622
…	..	..	..	..	..	…	..
Sim(LA_c, TA)						0.793711383	

Table 1: An example of a learner's score calculation

3.4 The Multidimensional Assessment Algorithm

This section illustrates the overall algorithm for the multidimensional assessment method as shown in the Algorithm: *The Multidimensional Assessment*. This algorithm delineates the steps to assess and calculate the final score for each learner in a multidimensional assessment scheme, as described below:

Lines 1 to 6: These steps prepare the assessment setting environment.

Lines 7 to 10: These steps perform the automatic assessment process on free-text answers and calculate $Score2_j$ for each $learner_j$.

Lines 11 to 17: These steps perform the M-DA between learners. We apply Formula (6) to calculate $Score3_{i,j}$ of $learner_i$ assessing $learner_j$'s answer. When i=j and j=i, $Score3_{j,i}$ is also calculated.

Lines 18 to 22: These steps sort $Score3_{i,j}$ in descending order.

Lines 23 to 29: These steps select the maximum *m* scores of $Score3_{i,j}$ and the sum for these *m* scores.

Algorithm: The Multidimensional Assessment

Input: Learners' answers, teacher's answer and results of learners assessing each other.

Output: Assessment results of learners.

Method:

1. Create question and answer;
2. Set time for learners to answer question;
3. Organize learners to answer a question;
4. Set time for learners to assess others' answers;
5. Let $Score1_{i,j}$ be scores of $learner_i$ assessing $learner_j$*'s* answer i, j = 1..n, i ≠ j;
6. Let $Comment1_{i,j}$ be comments of $learner_i$ assessing $learner_j$*'s* answer;
7. For each LA_j Do
8. Perform the assessment process on a free text answer;
//Automatic assessment process is proposed in section 3.3.
9. Let $Score2_j$ be the score of the system assessing $learner_j$*'s* answer, j = 1..n;
10. Endfor;
11. For each $Score2_j$ of $learner_j$ Do
12. For each $Score1_{i,j}$ of $learner_i$ assessing $learner_j$*'s* answer (i ≠ j) Do
13. System calculates $Score3_{i,j}$ of $learner_i$ assessing $learner_j$*'s* answer via formula (6);
//$Score3_{i,j}$: score of the system assessing the result when $learner_i$ assesses
//$learner_j$*'s* answer
//if i=j and j=i, then $Score3_{j,i}$ is also calculated.

14. Endfor;
15. Endfor;
// Sort the $Score3_{i,j}$ values in descending order
16. For every $learner_i$ Do
17. For every $learner_j$ that is assessed by $learner_i$ Do
28. Sort $Score3_{i,j}$ in descending order;
29. Endfor;
20. Endfor;
21. $Learner_i.Sum\_Score\_3 = 0$;
22. For every $learner_i$ Do
23. For $learner_j$ is assessed by $learner_i$ Do
//We choose the top of *m Scores3* of $learner_j$
24. $Learner_i.Sum\_Score\_3 = Learner_i.\ Sum\_Score\_3 + Score3_{i,j}$;
25. Endfor;
26. Endfor;
27. For every $learner_i$ Do
28. $Learner_i.final\_score = Score2_i * \alpha + ((Sum\_Score\_3 \text{ of } learner_i)/m) * \beta$;
29. Endfor;
30. For every $learner_i$ Do
31. Display(Final result of $learner_i$);
32. Endfor;
33. Return.

The following formulas are used in the multidimensional algorithm:

- The Euclidean distance[4]

Lines 30 to 32: These steps calculate the final scores for each learner and report to each learner. The final scores are calculated by employing Formula (8). For these steps, if a $learner_c \in learner_i$ assesses other learners with scores that are related or close to the system score, then $learner_c$ will obtain a high final score. This process aims to encourage each learner to actively assess other learners and to give a chance for each learner to discuss the material with other learners.

To calculate the value of $Score3_{i,j}$ for each $learner_i$ assessing $learner_j$*'s* answer, we have applied the Euclidean distance and marking technique to generate the $Score3_{i,j}$ formula, as shown in Formula (5).

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (5)$$

where D is the Euclidean distance between two points (x_1, y_1) and (x_2, y_2).

- $Score3_{i,j}$ is calculated via Formula (6), as shown below:

$$Score3_{i,j} = Max - \sqrt{(Score2_j - Score1_{i,j})^2} \qquad (6)$$

where $Score3_{i,j}$ is the score of the multidimensional assessment for each learner and *Max* is the maximum score in the marking scheme. In this study, *Max* = 10 (0=<score<=10). $Score2_j$ is the score of the system assessing $learner_j$*'s* answer, and $Score1_{i,j}$ is the score of $learner_i$ assessing $learner_j$*'s* answer, $i, j = \overline{1,n}$, $i \neq j$.

- The final score of $learner_i$ ($FSc\_learner_i$) is computed as follows:

$$\mathbf{FSc\_learner_i = (Score2_i) * \alpha + (Sum(Score3_{i,j})/m) * \beta;} \qquad (7)$$

where $Score3_{i,j}$ is calculated via Formula (6). Here, α and β are coefficients or weight values of the system; these variables assess the learners' answers and the results obtained from learners assessing other learners' answers, respectively. For this study, we set the values of α and β to be 0.7 and 0.3, respectively.

To decrease the difficulty of learners assessing each other, we set the system so that one learner is allowed to assess *m* other learners or more on a single question, and we select the maximum *m* scores of $Score3_{i,j}$ to calculate the final score. In our work, we set parameter *m* equal to 5. Therefore, Formula **(7)** becomes

$$\mathbf{FSc\_learner_i = (Score2_i) * 0.7 + (Sum(Score3_{i,j})/5) * 0.3} \qquad (8)$$

---

[4] http://mathworld.wolfram.com/Distance.html

Formula (8) is used to calculate the final score based on $Score2_i$, $Score3_{i,j}$ and their important weights.

Question1:	What is an operating system? *The answer criteria design: {Approximate size: approximately 80 words}, {answer order: definition, functions}*
Teacher's Answer:	An operating system is a software program or a set of programs that mediate access between physical devices and application programs.

Table 2: Example of question and teacher's answer

Learner$_j$' answers	**Learner$_i$ assesses Learner$_j$ with Score1$_{i,j}$ and Comment$_{i,j}$) (i≠j)**	**System assessment**		**Display the last Scores**
		Learners' answers (Score2$_j$)	**Results of learners assessing each other (Score3$_{i,j}$)**	
Learner$_{20}$'s answer	Learner$_{19}$ assesses Learner$_{20}$ with $Score1_{19,20}$: 9.0 Comment$_{19,20}$: …. - Correct information: - Deficient information: $Score1_{3,20}$: 4.0 score $Score1_{4,20}$: 8.0 $Score1_{5,20}$: 1.0 $Score1_{6,20}$: 6.0	Score2$_{20}$ of Learner20: **9.0**	$Score3_{19,20}$ of Learner 19:10 $Score3_{3,20}$: 5.0 $Score3_{4,20}$: 9.0 $Score3_{5,20}$: 2.0 $Score3_{6,20}$: 7.0	Final Score of learner20 = 9.0*0.7+ (Sum(9.0, 10, 9.3, 7.7, 9.0)/5)*0.3 = **9.0**
Learner$_{19}$'s answer	$Score1_{20,19}$: 8.5	$Score2_{19}$:9.5	$Score3_{20,19}$: **9.0**	9.1
	$Score1_{3,19}$: 4.5		$Score3_{3,19}$: 5.0	
	$Score1_{4,19}$: 8.5		$Score3_{4,19}$: 9.0	
	$Score1_{5,19}$: 1.0		$Score3_{5,19}$: 1.5	
	$Score1_{6,19}$: 7.5		$Score3_{6,19}$: 8.0	
Learner$_3$'s answer	$Score1_{20,3}$: 4.5	$Score2_3$:4.5	$Score3_{20,3}$: **10**	4.5
	$Score1_{19,3}$: 3.5		$Score3_{19,3}$: 9.0	
	$Score1_{4,3}$: 7.5		$Score3_{4,3}$: 7.0	
	$Score1_{5,3}$: 9.5		$Score3_{5,3}$: 5.0	
	$Score1_{6,3}$: 8.5		$Score3_{6,3}$: 6.0	
Learner$_4$'s answer	$Score1_{20,4}$: 9.0	$Score2_4$:8.3	$Score3_{20,4}$: **9.3**	8.5
	$Score1_{19,4}$: 6.0		$Score3_{19,4}$: 7.7	
	$Score1_{3,4}$: 4.5		$Score3_{3,4}$: 6.2	
	$Score1_{5,4}$: 0.5		$Score3_{5,4}$: 2.2	
	$Score1_{6,4}$: 7.5		$Score3_{6,4}$: 9.2	
Learner$_5$'s answer	$Score1_{20,5}$: 6.0	$Score2_5$:3.7	$Score3_{20,5}$: **7.7**	3.5
	$Score1_{19,5}$: 2.5		$Score3_{19,5}$: 8.8	
	$Score1_{3,5}$: 8.0		$Score3_{3,5}$: 5.7	
	$Score1_{4,5}$: 3.0		$Score3_{4,5}$: 9.3	
	$Score1_{6,5}$: 7.5		$Score3_{6,5}$: 6.2	
Learner$_6$'s answer	$Score1_{20,6}$: 6.5	$Score2_6$:5.5	$Score3_{20,6}$: **9.0**	6.0
	$Score1_{19,6}$: 5.5		$Score3_{19,6}$: 10	
	$Score1_{3,6}$: 3.5		$Score3_{3,6}$: 8.0	
	$Score1_{4,6}$: 5.5		$Score3_{4,6}$: 10	
	$Score1_{5,6}$: 0.5		$Score3_{5,6}$: 5.0	

Table 3: Example of the multidimensional assessment results

In *Tables 2* and *3*, we illustrate a concrete example by employing the multidimensional assessment algorithm. *Table 2* provides a question and an answer criteria design as given in the first row, and the teacher's answer is given in the second row. For *Table 3*, the first column contains several learners' answers. The second column contains the results of each learner assessing the others' answers, which include both the score and a comment; the comment includes both correct and deficient information. The system automatically assessed learners' answers and the results of learners assessing each other, and it then provided the results in Columns 3 and 4. The last column contains the final learners' scores.

The experiment and results are described in detail in the next section.

4 Experiment and Results

An experiment was conducted to assess learners in an e-course *"Introduction to Computer Science"* of Moodle LMS. The e-course was designed with fifteen open-ended questions and fifteen answers for the online learning environment. The experiment aims to enhance the efficient learning of learners learning the e-course. Ten questions were used to assess learners while they are learning, and five questions were used to assess learners when the e-course was completed. The class was divided into two groups with the same level of knowledge and skill, and each group contained twenty learners.

To evaluate the proposed method, this study developed two systems in Java, i.e., System1 and System2, and employed the MySQL server 5.1/MySQL Workbench 5.2 CE and NetBeans IDE 7.1.2 for a system implementation environment. Each system was applied to assess each group of students, namely,

- *System 1* is a multidimensional assessment system that was designed to assess the free-text answers of learners in *group 1*.

- *System 2* is an assessment system that was designed for assessing free-text answers. The same techniques as described in Section 3.3 was used but without the M-DA technique. This system was used to assess learners in *group 2*.

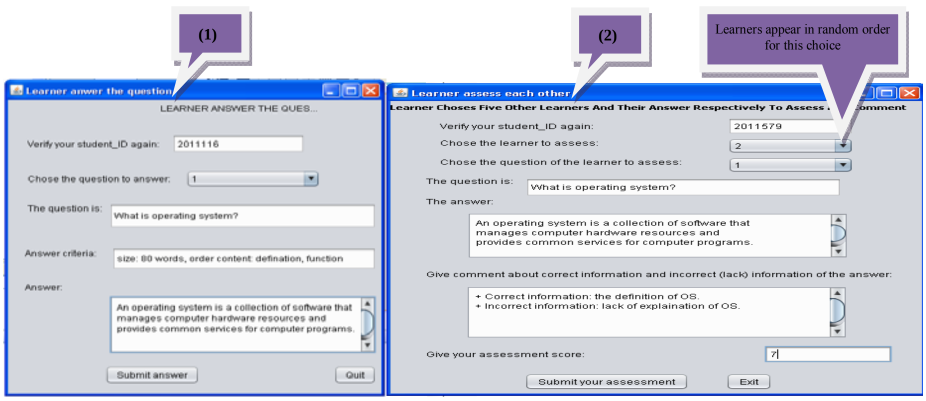

Figure 5: Screenshot from some learner activities in System 1

For this research, the dataset was composed of 600 answers of two groups. Learners assessed the answers of other learners 1,591 times by giving scores and comments in *group 1*, which were synthesized in *Table 4*.

Figure 5 contains some screenshots of *System 1*. Learners have two possible activities after a successful login. In the *test process* (1) of *Figure 5*, they can answer each question, and in the *assessment learner process* (2), each learner is allowed to choose other learners for assessing their answers.

The experiment results were analyzed and evaluated concretely as described below:

4.1 Evaluation for System 1:

In *System 1*, we collect the number of times that the learners assess others in fifteen questions of *group 1* students, as shown in *Table 4*.

Learners' order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Number of assessments	65	79	73	78	75	77	79	74	78	73	79	84	79	79	83	82	81	85	93	95

Table 4: The number of times the learners assessed others over fifteen questions

Once the e-course was finished, the final scores for each learner in *System 1* were calculated; the results are shown in *Table 5*.

Learners' order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Final scores	3.7	4.8	5.2	5.9	6.5	6.5	7.0	7.0	7.2	7.2	7.3	7.6	7.9	8.0	8.2	8.5	9.0	9.2	9.5	9.5

Table 5: The final results for each learner

In *Table 5*, the average score of the learners = 7.29 (where 0<= score <= 10).

In *Tables 4* and *5*, learner 1 assessed others 65 times, and his assessments were not exact because his score in *Table 5* is too low. Therefore, it can be concluded that he is rather inactive and lacks knowledge of the learning content. Conversely, learner 20 has proven that he is an active learner and has more knowledge in the learning content because he has given correct answers and received high scores for both the system score for his answers and the system's scores for assessing other learners' results.

The automatic multidimensional assessment method encourages learners to assess and share knowledge together because their assessment results were reassessed by the system. Therefore, learners usually interact and share comments with each other during the lesson.

4.2 Evaluation for System 2:

In *System 2,* the students in group 2 can only participate in studying and answering open-ended questions, and they are not allowed to assess other learners. The score of each learner for each question is calculated via *Score 2* using the one-dimensional assessment (the system assesses the learner answer only). The final scores for each learner are shown in *Table 6*, and are the average scores for fifteen questions.

Learners' order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Final scores	3.3	5.0	5.6	5.7	6.5	6.5	6.7	6.8	7.0	7.0	7.0	7.1	7.5	8.0	7.8	8.0	8.0	8.1	8.2	9.1

Table 6: The learners' final results

Table 6 indicates that the average score of all learners was 6.93.

The experiment results from *Tables 5* and *6* showed that the final results of learners in *group 1* were improved, with an average score of 7.29 (*Table 5*), whereas the average score of learners in *group 2* was 6.93 (*Table 6*). Therefore, the average score of learners in *System 1* increased 3.6% compared with *System 2*. Because the learners in the two groups had the same level of knowledge and skill, the learners in group 1 could spend more time actively assessing and commenting on their peers' answers. Hence, they could enhance the knowledge obtained from the e-course. We can conclude that the multidimensional assessment approach encourages learners to interact and collaborate actively and that this approach is more efficient than the single-dimensional assessment. *Figure 6* illustrates the comparison of the final scores between the systems.

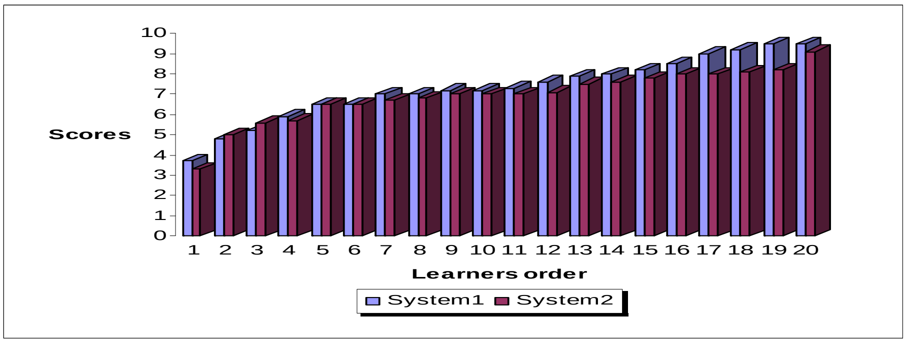

Figure 6: Comparison of the final results for the two systems

This research correlates with other research studies on open-ended question assessment techniques. The experiment results indicate that the efficiency of E-learning systems can be enhanced if the automatic multidimensional assessment is integrated into learning management systems.

5 Conclusions and Future Work

This research proposed a multidimensional assessment method that uses the M-DA system to assess free-text answers to enhance the efficiency of traditional e-learning systems; this approach enables interactions and collaborations in E-learning environments. The M-DA method aims to assess learners based on two criteria, i.e., the learners' knowledge comprehension and activities. Learners can answer the teacher's questions using the designated answer criteria without creating many reference answers. Each learner is allowed to assess other peers' answers by giving scores and comments. The M-DA system can evaluate the comprehension level of each learner using both system scores that are calculated for each learner and the scores obtained from learners assessing other learners. The evaluation results of the multidimensional assessment method increased by 3.6% compared with the approach without multidimensional assessment. Hence, the proposed method can enhance the interaction and collaboration of learners in a virtual learning environment.

This M-DA approach can be integrated with traditional learning systems to enhance the efficiency of the test evaluations, which is usually based on some types of questions, including multiple choices, true/false, short answers, and matching. The M-DA approach not only encourages learners to interact and collaborate actively but also provides learners with the opportunity to gain more knowledge from the comments of other learners through these activities.

In this research, we employ the vector space model and semantic extraction for assessing the free-text answers. However, the assessment technique relies on word comparisons. To enhance the efficiency of the assessment process, this work can employ other techniques, such as Semantic Web Technology, to enable the assessment of free-text answers in a more semantic manner. The assessment results should be more accurate and significant in the future.

Acknowledgments

This work was supported by a scholarship from the Khon Kaen University, Thailand, in the Ph.D. Program, Grant No. KKU Ref. 0514.1.11.1/5658.

References

[Abdalgader, 10] Abdalgader, K., Skabar, A.: Short-text similarity measurement using word sense disambiguation and synonym expansion, *Australasian Conference on Artificial Intelligence,* 2010, pp. 435-444.

[Alfonseca, 04] Alfonseca, E., A., Pérez, D.: Automatic assessment of open ended questions with a BLEU-inspired algorithm and shallow NLP, in *Proceedings of ESTAL*, 2004, pp. 25-35.

[Alier, 12] Alier, M., Mayol, E., Casañ, M.J., Piguillem, J., Merriman, J.W., Conde, M.Á., García-Peñalvo, F.J., Tebbens, W., Severance, C.: Clustering Projects for eLearning Interoperability, *Journal of Universal Computer Science*, Vol 18, no. 1, 2012, pp: 106-122.

[Assareh, 11] Assareh, A., Hosseini Bidokht, M.: Barriers to e-teaching and e-learning, in *Procedia Computer Science,* Vol 3, 2011, pp. 791-795.

[Caballé, 12] Caballé, S., C., Gañán, D., Dunwell, I., Pierri, A., Daradoumis, T.: CC-LO: Embedding Interactivity, Challenge and Empowerment into Collaborative Learning Sessions, *Journal of Universal Computer Science*, Vol 18, no. 1, 2012, pp: 25-43.

[Castellanos-Nieves, 11] Castellanos-Nieves, D., Fernández-Breis, J.T., Valencia-García, R., Martínez-Béjar, R., Iniesta-Moreno, M.: Semantic Web Technologies for supporting learning assessment, *Journal of Information Sciences,* Vol 181, 2011, pp. 1517-1537.

[Fardoun, 09] Fardoun, H., Montero, F., Jaquero V.L.: eLearniXML: Towards a model-based approach for the development of e-Learning systems considering quality, *Journal of Advances in Engineering Software,* Vol 40, 2009, pp: 1297–1305.

[Fardoun, 12] Fardoun, H., Alghazzawi, Daniyal M.: A Comparison of eLearning XML with current e-Learning System Development Methodologies, *International Conference on Information Human Computer Interaction & Learning Technologies,* 2012.

[He, 09] He, Y., Hui, S.C., Quan, T.T.: Automatic summary assessment for intelligent tutoring systems, *Journal of Computers & Education*, Vol 53, 2009, pp. 890–899.

[Hou, 10] Hou, W.J., Tsao, J.H., Li, S.Y., Chen, L.: Automatic assessment of students' free-text answers with support vector machines, in Proceedings of IEA/AIE, Vol 1, 2010, pp. 235-243.

[Hurtado, 11] Hurtado, C., Guerrero, L. A.: Enhancement of Collaborative Learning Activities using Portable Devices in the Classroom, *Journal of Universal Computer Science*, Vol 17, no. 2, 2011, pp: 332-347.

[Islama, 11] Islama Md. S., Kunifuji, S., Hayama, T., Miura, M.: Towards exploring a global scenario of e-learning in library and information science schools*, Journal of International Information & Library Review,* Vol 43, 2011, pp: 15-22.

[Liccardi, 07] Liccardi, I., Ounnas, A.., Pau, R., Massey, E., Kinnunen, P., Lewthwaite, S., Midy, M.A., Sarkar, C.: The role of social networks in students’ learning experiences, ACM SIGCSE Bulletin (December Issue), 2007, pp. 224-237.

[Loc, 12] Loc, H.P., Ngamnij, A., Somjit, A., Wararat, R.: Multi-dimensional assessment on free-text answers to enhance learners’ activities and collaborations, in *Proceedings of the 2012 IEEE 16th International Conference on CSCWD,* 2012, pp.153-159.

[Noorbehbahani, 11] Noorbehbahani, F., Kardan, A.A.: The automatic assessment of free text answers using a modified BLEU algorithm, *Journal of Computers & Education,* Vol 56, 2011, pp. 337–345.

[Ozkan, 09] Ozkan, S., Koseler, R.: Multi-dimensional students’ evaluation of E-learning systems in the higher education context: An empirical investigation, *Journal of Computers & Education,* Vol 53, 2009, pp. 1285–1296.

[Tissenbaum, 12] Tissenbaum, M., Lui, M., Slotta, J.D.: Co-Designing Collaborative Smart Classroom Curriculum for Secondary School Science, *Journal of Universal Computer Science*, Vol. 18, no. 3, 2012, pp. 327-352.

[Wang, 10] Wang. L.: How social network position relates to knowledge building in online learning communities, *front. Education China,* Vol 5(1), 2010, pp. 4–25.

[Wong, 07] Wong, D.: A critical literature review on e-learning limitations, in *School of Management & Information Technology*, JASA 2, January 2007.

[Zhang, 08] Zhang, W., Yoshida, T., Tang, X.: Text classification based on multi-word with support vector machine, *Journal of Knowledge-Based Systems,* Vol 21, 2008, pp. 879–886.