Prototype of Automatic Essay Assessment and Plagiarism Detection on Mobile Learning "Molearn" Application Using GLSA Method

Julianto Lemantara
Department of Information System
Dinamika University
Surabaya, Indonesia
julianto@dinamika.ac.id

M.J. Dewiyani Sunarto
Department of Information System
Dinamika University
Surabaya, Indonesia
dewiyani@dinamika.ac.id

Bambang Hariadi
Department of Film & Television Production
Dinamika University
Surabaya, Indonesia
bambang@dinamika.ac.id

Tri Sagirani
Department of Information System
Dinamika University
Surabaya, Indonesia
tris@dinamika.ac.id

Tan Amelia
Department of Information System
Dinamika University
Surabaya, Indonesia
meli@dinamika.ac.id

***Abstract*— In evaluating the student's learning outcomes, essay exams were commonly used by teachers to measure the level of student's understanding of the learning material. However assessing essay answers was more difficult in reality because it contained teacher's subjectivity and required a longer correction time. In addition, detecting similarity in essay answers between students also required more teacher's efforts. In previous studies, a prototype of essay answer assessment and plagiarism detection had been successfully created. However, the prototype display still needed an improvement based on the evaluation results given by biology teachers in East Java Province as the application users. The previous prototype also still carried the Latent Semantic Analysis (LSA) method which had several weaknesses. Therefore, this study aimed to produce prototypes that had better display and text similarity methods. The Generalized Latent Semantic Analysis (GLSA) method was chosen because it was able to cover the weaknesses of the LSA method. GLSA was able to detect sentences that had syntactic errors or missing common words. Based on the evaluation results, this study succeeded in producing a prototype with a better display value. The level of user satisfaction increased by 6.12%. In addition, the study succeeded in using the GLSA method as a substitute for LSA for creating better prototype essay assessment and automatic plagiarism detection.**

Keywords*— *mobile learning; molearn; plagiarism; automatic essay assessment; generalized latent semantic analysis

I. INTRODUCTION

In education field, the process of evaluating student's learning outcomes comes to be an important component as the result indicate students' understanding towards the taught materials. The evaluation process could be performed through some assessment types using different question formats such as multiple choice and essay. When the multiple choice offers some optional answers, the essay format let the students have their own sentences, of which make them more trained in verbal communication [1]. The ability of argumentation needs to be exercised due to building theoretical understanding of a concept. This ability directs the students to have perceived clear and new knowledge [2]. Similarly the essay test requires a better understanding of a knowledge and can be used to measure the level of human understanding of a knowledge in depth [3]. The Head of Research and Development (Balitbang) of the Ministry of Education and Culture states that the presence of essay test encourages students to ably argue, give reasons, and solve problems. He further explained that the notion of education should be relied on the objective of creating critical students, for instance, through the exercise of essay test in which students are exposed more into logical thinking. Therefore, essay questions are mandatory in any examination as simply it trains students to conceive more exposures on critical thinking skills.

The essay-type question is usually preferred by teacher in evaluating students' understanding level eventhough providing its objective assessment still remains clueless. Even there is an assessment rubric for essay question, the emergence of subjectivity still exists. The teacher also needs extra working hours just to check the essay answers that promote inefficiency. This might influence the decrease of assessment quality as the teacher is likely confronting exhaustion that leads to random judgments [4]. Henceforth, an automatic tool

for assessing essay answers should be invented in order to overcome some above dilemma.

Many previous studies have been successful in developing a website and mobile-based learning application called Mobile Learning (Molearn). Website-based applications can be accessed at http://molearn.net while the mobile-based application can be downloaded via Google Play store. Molearn application is an electronic learning product (e-learning) that has been implemented in several schools of East Java regions. The use of e-learning seems familiar today as most educational institutions use it to deliver the materials, lecturing process, or even assessment process, of which all features are equipped with remote access networks [1][5].

Molearn can assist teacher to conduct evaluation of learning outcomes. However, the application is still limited on carrying out an evaluation for multiple choice question types and not the essay ones. Due to the unavailability of the application, 97.22% of biology teachers in East Java Province agreed the use of an automatic essay assessment feature because, in the same time, they also easily detected the similar essay answers or checked plagiarism level. However, they still need extra efforts to reveal those who cheat during the exam.

In connection with teacher's complaints who use the Molearn application, a study was conducted to make a prototype of an online exam on the Molearn application using text similarity [6]. The aim was to provide an initial overview of the form of automatic essay assessment and check students' answers plagiarism on the Molearn application. However, the study only reached the prototype manufacturing stage, not up to the implementation stage. In addition, the prototype still used the LSA method for text similarity. Ruslan, et al conveys that several techniques used in automatically evaluating essays include Latent Semantic Analysis (LSA), Probalistic Latent Semantic Analysis (PLSA), and Generalized Latent Semantic Analysis (GLSA) [7][8]. In LSA, words are represented in a semantic matrix and then mathematically processed using Singular Value Decomposition (SVD) linear algebraic techniques. This method is relatively simple but has a fairly high level of correlation when compared to human-made assessments manually. The LSA method only assesses the similarity between text documents through the frequency terms that exist in each text document so that it has the disadvantage of not paying attention to the order of words or the layout of the terms which indirectly influences the meaning contained in each document. GLSA is a further development of the LSA algorithm by making n-grams based on a document matrix [7][9]. GLSA considers the order of the sentence words in the document and maintains the closeness of the words in the sentence. In general, GLSA fixes the lacks of LSA in terms of detecting sentences that have syntactic errors or missing common words. GLSA is proven to have a good accuracy value in evaluating essay answers automatically so that its method is carried out in this present study [7]. The difference between this study and Ruslan et al. is that Ruslan et al. only limits the study on matching students' essay answers with teacher's answer keys, but has not yet considered the level of plagiarism of essay answers among students. So the novelty of this study is the use of GLSA. In addition, the GLSA method is not only applied to automatically assess essay answers based on key reference answers, but it is applied to detect the level of similarity or plagiarism of students' essay answers.

Henceforth, the present study aims to produce a prototype of automatic essay assessment and plagiarism detection using GLSA method. This prototype is the initial form of the innovation of Molearn application. If this prototype is successfully implemented, the teacher does not need to correct students' examination answers one by one to save times and energy as well as to avoid plagiarism. In addition, the teacher can also get information if there are similar answers of students' essays so that the teacher can take decisive action in the event of fraud. Molearn application can provide learning evaluation strategies that are appropriate to the learning needs of the 21st century and the Industrial Revolution 4.0 which require literacy, critical thinking, scientific creativity, collaboration, Information and Communication Technology, and problem solving skills [10][11][12].

II. METHODOLOGY

A. *Latent Semantic Analysis (LSA)*

LSA was a method that had a characteristic to extract and represent sentences with mathematical calculations and attach important key words contained in a sentence regardless the linguistic characteristics [7]. Mathematical calculations were performed by mapping the presence of words from groups of words on the semantic matrix and then processed using Singular Value Decomposition (SVD) linear algebra techniques. LSA was used to assess essay by converting it into matrices that are rated on each term to look for similarities with the reference terms.

Numbers of LSA steps in assessing essays encompassed Term Document Matrix that represented the teacher's and students' answers as a document matrix. This matrix consisted of rows and columns where rows represented a unique word from the teacher's overall answer while columns represented students' answers documents where the value of each row and column was the frequency in which the terms appeared in the document. Furthermore, the matrix decomposition process used SVD as well as the dimension reduction process aimed at reducing the magnitude of the document matrix produced and discarding data that did not have a strong correlation with the teacher's answer document data. Each word in the paragraph was represented as a row and column matrix. By using the SVD algebraic matrix technique, the matrix was decomposed into three matrix components namely two orthogonal matrices and one singular diagonal matrix. SVD was a linear algebra theorem which said that the rectangle matrix A could be broken down into three matrices namely:

a) Orthogonal U matrix
b) Diagonal S matrix
c) Transpose of the orthogonal V matrix.

The three matrices could be formulated as in the following equation 1:

$$A_{mn} = U_{mm} \times S_{mn} \times V_{nn}^{T} \quad (1)$$

Notes:

A_{mn} = initial matrix

U_{mm} = orthogonal U matrix

S_{mn} = diagonal S matrix

V_{nn}^{T} = transpose orthogonal V matrix

The final stage of the LSA process was to calculate the similarity in semantic terms using Cosine Similarity. Cosine Similarity was used to calculate the cosine value of an angle between two vectors (e.g. document and query vectors; the smaller the angle produced, the higher the level of similarity of an essay). Cosine Similarity could be formulated as in the following equation 2:

$$cos \propto = \frac{A.B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (2)$$

A = Document vector

B = Query vector

$A . B$ = dot multiplication Vector A and Vector B

$|A|$ = length of vector A

$|B|$ = length of vector B

$|A||B|$ = cross product between |A| and |B|

$\propto$ = the angle formed between Vectors A and B

Furthermore, Fig. 1 shows the details of the LSA stages.

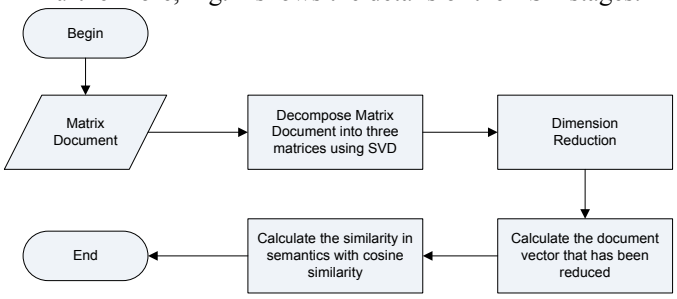

Fig. 1. The LSA Stages

B. Generalized Latent Semantic Analysis (GLSA)

GLSA was a further development of the LSA algorithm by making n-grams based on a document matrix [7]. GLSA considered the order of the sentence words in the document and maintained the closeness of the words in the sentence. It was not based on document vectors with the bag of words approach. The process began with the similarities between words paired semantically to compute representations for words.

The traditional word by document matrix creation of LSA does not consider word sequence in a document [13][14]. In the word formation by document matrix, the word pair "concurrent transactions" produces the same result of "transactions concurrent". Therefore, LSA fails to capture the sematic effect of collocations in the document. GLSA resolves this problem by considering n-gram as atomic unit of the document instead of individual word. So, in the GLSA method, "concurrent transactions" is not recognized same as "transactions concurrent"

In conclusion, eventhough GLSA process had similarities with LSA ones, however, the one that distinguished GLSA from LSA only at the stage of formation of N-gram documents that would focus on word order and maintain the closeness of words in sentences. N-gram was a long sequence of symbols written as unigram, bigram, trigram, and n-gram [15]. N-gram modeling, usually defined as the Markov chain sequence, was a model that maps the probability distribution of n-letter sequences. The N-gram model was widely used in language processing or called Natural Language Processing (NLP). In the current study, n-gram was used as an index. Therefore, N-Gram denoted a combination of words in a row, maintained the closeness of the sentence words in the document, and represented a phrase of size N.

C. SDLC Waterfall Model

In developing the application, this study used System Development Life Cycle (SDLC) waterfall model in which its stages covered:

1. *Requirement Analysis*, was to look for data requirements, user needs, and functional and non-functional requirements of the application. The workflow and functional requirements of the prototype are still the same as previous research conducted by the research team itself [6]. For more details, the workflow and functional requirements of the prototype can be seen at Fig. 2 dan Fig. 3. Fig. 2 shows there are two main users in the online exam system. They are teachers and students. Teachers have a job to make questions and answer keys. In this research, questions can be either a multiple choice or essay. Besides, teachers have a job to review the essay assessment results. On the other hand, students have two jobs: answer the exam questions and view the exam scores.

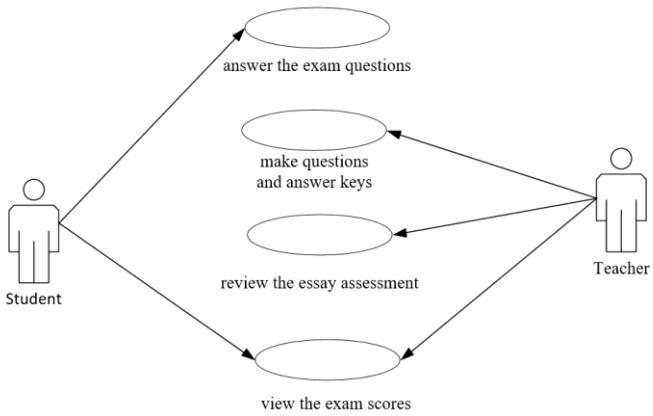

Fig. 2. The Functional Requirements

Basically, this prototype requires input in the form of questions and answer keys coupled with students' answers during the exam. Afterwards, the system will generate output in the form of multiple choice test scores, score recommendation, and plagiarism level of the students' answers. Fig. 3 shows how this prototype works.

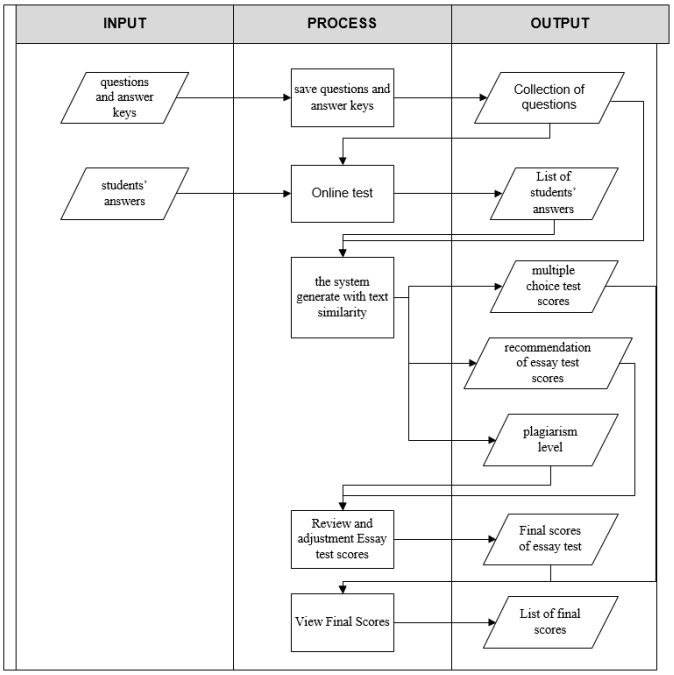

Fig. 3. The Workflow of the prototype

2. *Design*, was to design a system starting from designing a database, user interface, and input / output design.
3. *Development*, was the stage of application-making process. In this case, the application was an automatic essay assessment using GLSA. The GLSA stages comprised making matrix documents, decomposing matrix documents using SVD, dimension reduction, vector calculation, and semantic similarity calculation with cosine similarity.
4. Testing, was the stage of testing the new prototype. After trying the prototype, the teachers filled the questionnaire to evaluate the prototype. In this stage, the level of user satisfaction could be known. This research compared with previous research regarding the value of the prototype display.
5. *Maintenance*, was the stage of implementing and maintaining the application quality. This stage was not conducted in this research because the application has not been produced. This research is only limited to a prototype. So, it is still in the form of an initial design.

Fig. 4 depicts the details of SDLC with the waterfall model.

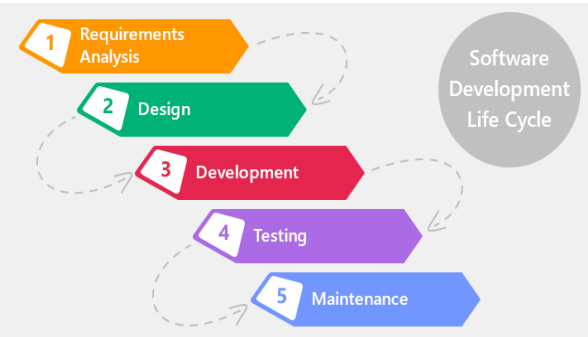

Fig. 4. SDLC Waterfall Model

III. Results And Discussion

Several previous studies had produced the Molearn application with the ability to automatically correct multiple choice answers. This was easier to do for student's answers are only compared to the answer key. Fig. 5 shows the user interface of automatic assesment for multiple choice answers.

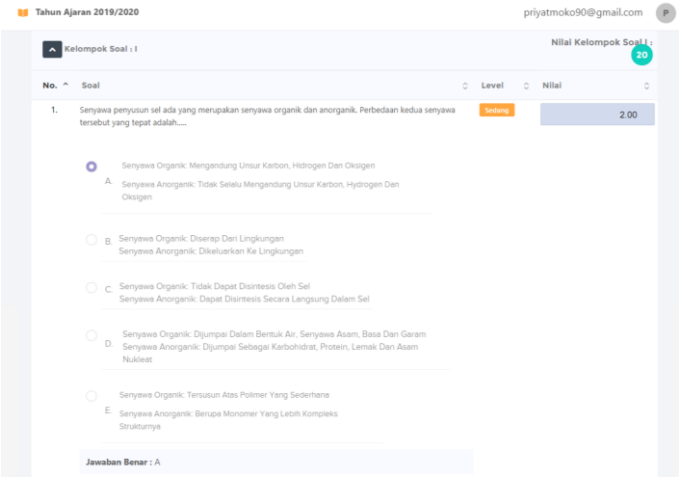

Fig. 5. The user interface of automatic assesment for multiple choice answers.

The evaluation of essay answers was clearly different from the assessment of multiple choice answers because the essay answers were typically open-responses. Students could answer freely and did not have to be exactly the same as the answer key. This was what made the computerized grading of the essay answers more difficult than that of multiple choice answers. Currently, a new study focused on making prototypes of assessment and detecting plagiarism levels of students' essay answers automatically. Fig. 6 portrays the prototype of the automatic essay assessment and plagiarism detection.

Fig. 6. The prototype of the automatic essay assessment and plagiarism detection.

Fig. 6 shows that the Molearn application could be further developed in accordance with the undertaken prototype design so that the application was more useful for teachers in evaluating learning outcomes. In Fig. 6, the teacher could immediately look at the recommended value of the students' essay answers to the answer key. In addition, the teacher could also see the highest level of plagiarism from one student's answer and compare it to another student's answer. In regard to the developed prototype design, the teacher could reveal the percentage value of the similarity of students' answers accompanied by the students' names who had the answers most similar to those being assessed. This prototype had been approved by several biology teachers who became the members of an organization called the East Java Biology

Teachers' Community (MGMP). This was evidenced by the fact that 35 out of 36 (97.22%) biology teachers who were the members of the MGMP supported the development of this prototype because it offered effective assistance to correct essay answers. The teacher's satisfactory level towards the prototype was also relatively high at 77.8%. From a scale of 1 (very dissatisfied) to 5 (very satisfied), two teachers gave a grade of 2, 6 teachers gave a grade of 3, 23 teachers gave a grade of 4, and 6 teachers gave a grade of 5. In total, the average grade obtained was (1*2 + 5*3 + 22*4 + 8*5) / (1 + 5 + 22 + 8) = 145 / 36 = 4.028. In connection with its percentage, the teacher's satisfactory level was 4.028 / 5 = 80.56%. Thus, this prototype was feasible to be applied subsequently into an application that was ready for use in various schools. Fig. 7. shows the questionnaire results of the new prototype.

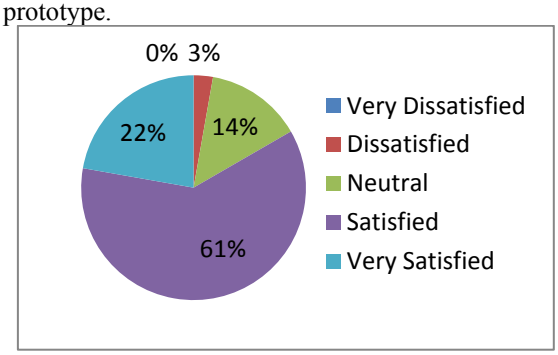

Fig. 7. The questionnaire results of the new prototype.

The prototype results had improved the previous prototype in terms of application display when assessing automatic essay responses and detecting plagiarism answers. This was evidenced by the results of the questionnaire which showed that the prototype in previous studies only received the values of (3*2 + 9*3 + 19*4 + 5*5) / (3 + 9 + 19 + 5) = 134 / 36 = 3.722. The satisfactory level of the previous prototype was 74.44%. Fig. 8 provides the questionnaire results of the previous prototype.

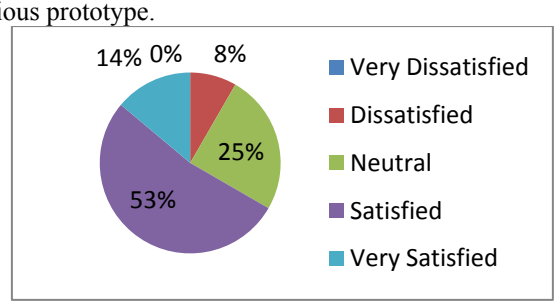

Fig. 8 The questionnaire results of the previous prototype

From the questionnaire results, the prototype display in this study had increased the satisfactory level. This was due to the fact that the prototype display of the present study in Fig. 6 separated the value of the suitability of the answer and the value of plagiarism to students' answers and the answer key. In addition, some teachers also gave reasons that the display of the latest prototype was simpler and neater. Fig. 9 shows the results of previous prototypes.

Fig 9. The previous prototype.

In addition to increasing the value of the display, the prototype developed in this study carried a better text similarity method than the previous prototype. If the previous prototype used the LSA method, this study applied the GLSA method. GLSA was actually the development of LSA so that GLSA could fix the weaknesses of LSA in terms of detecting sentences that had syntactic errors or missing common words. Several studies had also shown that GLSA was better than LSA. Thus, the prototype produced in this study had two strengths namely: a better display value based on its users' perspective and the use of sophisticated text similarity methods. In brief, the differences between current study and previous study can be explained in Table I.

TABLE I
THE DIFFERENCES BETWEEN TWO STUDIES

Factor	Previous Study	Current Study
Method of text similarity	Latent Semantic Analysis	Generalized Latent Semantic Analysis (better method)
Level of user satisfaction	74.44%	80.56% (better display in the new prototype)

IV. CONCLUSIONS AND RECOMMENDATIONS

Based on the above explanation, this study concludes that:

1. In connection with the biology teachers' responses, the prototype developed in this study has a better display compared to the previous prototype in showing the similarity of students 'answers with the answer key and the plagiarism level of students' answers. This was proven by an increase in the percentage of user satisfaction by 6.12%, from 74.44% to 80.56%.
2. The present prototype developed in this study uses GLSA method so that it can improve the previous prototype in detecting text similarity for automatic essay assessment and plagiarism detection.

ACKNOWLEDGMENT

The research was supported by Ministry of Research, Technology and Higher Education in the Republic of Indonesia. Besides, it was also supported by Institute of Business and Informatics Stikom Surabaya.

REFERENCES

[1] Ratna, A.A.P., Budiardjo, B., & Hartanto, D., "SIMPLE : Sistem Penilaian Esei Otomatis Untuk Menilai Ujian Dalam Bahasa Indonesia", *Makara, Teknologi*, Vol. 11, No. 1, pp 5-11, 2007.

[2] Wardani, A. D., Yuliati, L., & Taufiq, A., "Kemampuan Argumentasi Ilmiah dan Pemecahan Masalah Fisika Siswa SMA pada Materi Gaya

dan Gerak. Seminar Nasional Pendidikan IPA", Malang: *Pasca Sarjana UM*, pp. 13-28, 2016.

[3] Hamza, S., Sarosa, M., & Santoso, P. B., "Sistem Koreksi Soal Essay Otomatis Dengan Menggunakan Metode Rabin Karp", 2013.

[4] Fitri, R., & Asyikin, A.N., "Aplikasi Penilaian Ujian Essay Otomatis Menggunakan Metode Cosine Similarity", *Jurnal POROS TEKNIK*, Vol. 7, No. 2, pp 88-94, 2015.

[5] Aji, R. B., Baizal, A., & Firdaus, Y., "Automatic Essay Gradding System menggunakan Metode Latent Semantic Analysis", *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Yogyakarta, ISSN : 1907-5022, 2011.

[6] Lemantara, J., Sunarto, M.J.D., Hariadi, B., Sagirani, T., dan Amelia, T., "Prototype of Online Examination on MoLearn Applications Using Text Similarity to Detect Plagiarism", *5th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, Semarang, 2018.

[7] Ruslan, Gunawan, dan Tjandra, S., "Sistem Penilaian Otomatis Jawaban Esai Menggunakan Metode GLSA", *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, ISSN: 1907 – 5022, 2018.

[8] Olney, A.M., "Generalizing Latent Semantic Analysis," in *IEEE International Conference on Semantic Computing*. University of Memphis, Memphis, USA, 2009.

[9] Islam, M. M., dan Hoque, A. S. M. L., "Automatic esai scoring using Generalized Latent Semantic Analysis", *13th International Conference on Computer and Information Technology (ICCIT)*, 2010.

[10] Griffin, P., & Care, E., "Assesment and Teaching of 21 st Century Skills: Methods and Approach", New York: *Springer*. 2015.

[11] Jatmiko, B., Prahani, B. K., Munasir, Supardi, Z., Wicaksono, I., Erlina, N., et al., "The comparison of OR-IPA teaching model and problem based learning model effectiveness to improve critical thinking skills of pre-service physics teachers", *Journal of Baltic Science Education*, 17(2), pp 1-22, 2018.

[12] Pandiangan, P., Sanjaya, I. M., & Jatmiko, B., "The validity and effectiveness of physics independent learning model to improve physics problem solving and self-directed learning skills of students in open and distance education systems", *Journal of Baltic Science Education*, 16(5), pp 651-665, 2017.

[13] Guruji, P.A., Pagnis, M.M., Pawar, S.M., & Kulkarni, P.J., "Evaluation of Subjective Answers Using GLSA Enhanced with Contextual Synonymy", *International Journal on Natural Language Computing (IJNLC)*, Vol. 4, No. 1, February 2015.

[14] Devi, M.S., & Mittal, H., "Subjective Evaluation using LSA Technique," in *International Journal of Computers and Distributed Systems*, Vol. No.3, Issue I, ISSN: 2278-5183, April-May 2013.

[15] Russell, S. J., Norvig, P., dan Canny, J., "Artificial intelligence: a Modern Approach", 2003.