International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 1
ISSN: 2088-5334

Automatic Essay Assessment for Blended Learning in Elementary School

Endang Widi Winarni [a,*], Endina Putri Purwandari [a], Siti Hafiza [a]

[a] *University of Bengkulu, Bengkulu, 38371A, Indonesia*
Corresponding author: [*]endangwidi@unib.ac.id

***Abstract*—Blended learning combines traditional face-to-face learning in class and virtual learning. This system requires an evaluation process as a measuring instrument. This paper develops an automated essay correction on a blended learning system for elementary school. Assessment in education is obtaining, organizing, and presenting information about what and how the student learning. Open-answer questions allow teachers to understand the student's answer. Essay questions can be used to train students in conveying information verbally and measure their understanding. The teacher needs more time to examine the essay answers for each student. The essay correction needs to be guided with a scoring rubric as the keyword in the answer key that automatically makes essay corrections for elementary school. This system uses the Rabin Karp method to measure the similarity between answer keys to students' answers. The test was carried out by comparing Mean Absolute Error and Pearson Correlations from various *k*-gram values. The experiments show this assessment system produces a small error value and good performance in grading the student's answer with a low difference value between automatic assessment and expert judgment. Further research, this system can be applied to evaluate the student learning outcomes in an integrated manner with STEAM elements through blended learning. The use of automatic essay assessment in blended learning can improve elementary school students writing skills in the digital educational environment 4.0.**

***Keywords*— Automatic essay assessment; blended learning; elementary school; Rabin Karp.**

Manuscript received 28 Apr. 2020; revised 25 Feb. 2021; accepted 28 Mar. 2021. Date of publication 28 Feb. 2022.

I. INTRODUCTION

E-learning is a learning process supported by technological services such as telephone, audio, videotape, satellite transmission, or computer. E-learning is a learning activity that utilizes an internet network as a delivery method, interaction, and facilities supported by various forms of other learning services. An exam can be done online in blended learning, starting from answering exam questions to giving exam scores. This provides advantages for teachers and students because computer-assisted correction systems can provide faster and more accurate assessments than traditional assessment systems [1].

Blended learning combines traditional face-to-face learning in class and virtually (e-learning) using information technology through websites [2]. Blended learning methods enable students to access various learning resources based on text, images, videos, and animations. The material is presented briefly, densely, and systematically, and rich learning resources help students learn actively in a short time. The effectiveness of learning can increase motivation and student learning outcomes [3]. Blended or hybrid learning occurs to accommodate the learning strategies with Industrial Revolution 4.0 [4]. This model encourages educators to change the educational paradigm from teacher-centered learning to student-centered learning. Blended learning with interactive learning media in era education 4.0 gives some advantage [5] such as (1) students can learn independently without waiting for face-to-face learning in class; (2) fun and interesting learning through various instructional modalities or delivery media; (3) students can practice deepening competence and knowledge by combining instructional method; (4) interaction between teachers and students. Students can practice the exercises through blended learning. Then the system automatically evaluates the answer essays from students.

An essay exam is a test in structured questions, and students arrange, organize answers to each question in their language. Varied responses from students characterize the essay, not only selected from multiple choices. This question is a form of evaluation where answer choices are not provided, and students must answer with sentences. Hence, the answers vary significantly according to the thoughts of each examinee.

Online assessment in e-learning is expected to provide effective and efficient assessment methods for teachers and students [6].

Online testing is an evaluation process through computer and internet technology. The automated essay correction system evaluates learning outcomes conducted with computer technology and connected to the internet [7]. This system can contribute to time resources and humans to reduce the inequalities judgments between examinees and cause subjectivity scores obtained by examinees. The automatic essay correction system with Rabin Karp can be the right solution. Rabin Karp is a simple method with multipattern completeness. Essay correction is still challenged for marking and time-consuming assessments. It is difficult to define one correct answer because of the differences between experts. The essay correction needs to be guided with a scoring rubric as the keyword in the answer key. An expert might access the same essay to answer differently, depending on the answer is looked at known as marker drift. Ideally, the essay should be corrected by the other expert to enhance the marking reliability.

Some researchers develop automatic essay scoring research to solve the reliability problem and time-consuming assessment. Thomas [8] proposes an automated assessment system for descriptive student answers using Latent Semantic Analysis (LSA). The score is generated from correct keyword usage and spell-check. Liang [9] decompose model Siamese Bidirectional Long Short Term Memory Architecture (SBLSTMA) automatically scoring essays. This model shows better performance than neural network methods. Citawan [10] built automatic essay scoring with LSA and *n*-gram features. These experimental results show the unigram term gives the highest accuracy or lowest error rate between expert and system scores. Yudhana [11] used the Rabin Karp algorithm to compare the digital document's similarity in PDF file. Frinhani [12] developed the correction system in an active learning environment using LSA. Statistically, the score difference between system and expert are very close. Setiawan [13] developed a model in e-learning discussion using LSA. Statistically, the framework has a good performance to reveal the topic of the discussion forum. Kim [14] used e-portfolios as an evaluation activity to promote reflective thinking, critical thinking, and problem-solving ability. Anongnad Petchpraset [15] designed an automated writing evaluation tool named Coh-Metrix to grade the English as Foreign Language (EFL) for undergraduate students in Thailand. This automated text analysis help teacher in grading and improve the students writing skill.

Blended learning, which combines information technology activities with traditional face-to-face classrooms, gave an excellent opportunity to use the best assessment tools. Testing provides benefits for both students and teachers in traditional, blended, and learning contexts. This research aims to design an essay question application and grading system so that teachers can view test results, analyze student grades, and identify the difficulty level for each question. Next, develop automatic essay corrections on a blended learning system for elementary schools. This article can contribute to education assessment by organizing, evaluating, and presenting information about how students answer open essay questions. It helps teachers know the understanding of concepts and critical thinking skills in marking the student answers.

II. Materials and Method

The blended learning model can shift the learning principle dynamically from teacher-center learning to student center learning. According to Carman [16], there are five keys to implementing blended learning, namely: (1) Live Event, direct learning or face-to-face synchronous in the same time and place or at the same time but different places; (2) Self-Paced Learning, which combines independent learning that allows participants to learn anytime, anywhere online; (3) Collaboration, combining collaboration, both teacher and students collaboration; (4) Assessment, the designer must be able to formulate a combination of online and offline assessment types both test and non-test; and (5) Performance Support Materials, make sure learning materials are prepared in digital form, accessible to study participants both offline and online.

Rabin-Karp algorithm is a string search algorithm found by Michael O. Rabin and Richard M. Karp in 1987. This algorithm is the simplest string searching algorithm that uses hashing to find a substring in a text [17]. Hashing is a method that uses a hash function to convert a data type into several simple integers. Rabin-Karp algorithm aims to finds a pattern from the input text. This algorithm produces good time efficiency in detecting strings that have more than one pattern. Rabin Karp uses a hash function as a comparison between strings (m) with substring in the text(n) [18]. If the two's results are not the same, then the substring will shift to the right. The shift is carried out as much as (n-m) times. If two strings are equal, then the hash value must be the same. This problem can be solved by assigning several strings with the same hash value.

The key to Rabin Karp's algorithm is in selecting hash values. The automatic essay correction stages include the preprocessing stage, such as case folding, tokenizing, stemming. Rabin Karp algorithm consists of stages of determining gram, hashing, and similarity. Rabin Karp compared the hash values from input strings and substrings in the text. If it is the same, then a comparison will be made of the characters. If not the same, then the substring will shift to the right. The primary key of algorithm is an efficient calculation of substring hash value at the shift time.

K-gram is a series of terms with length *K*. *K*-gram is applied to generate words or characters. This *k*-gram method is used to extract *k* characters or letters from a continuously read word to the document's end. Similarity calculations from a set of words use the Dice-Sorenson Coefficient for the pair of words [19]. The similarity value is calculated by equation (1), where *S* for similarity value, *C* for the number of *k*-grams from two compared texts, *A* and *B* for the number of *k*-grams from each compared text.

$$S = \frac{2C}{A + B} \tag{1}$$

The automatic essay correction workflow is divided into several stages: student answer, preprocessing, text processing, and similarity (see Fig. 1). This system is built for students and lecturers in the learning process in elementary school. Students do the exam given by the lecturer. The lecturer

provides the essay question to students then automatically assessed by the blended learning system based on the teacher's answer key. In the first stage, the instructor inputted the question-and-answer keys into the system. Then, students chose the exam and answered the essay directly to blended learning. This answer will be stored in the database and processed by preprocessing.

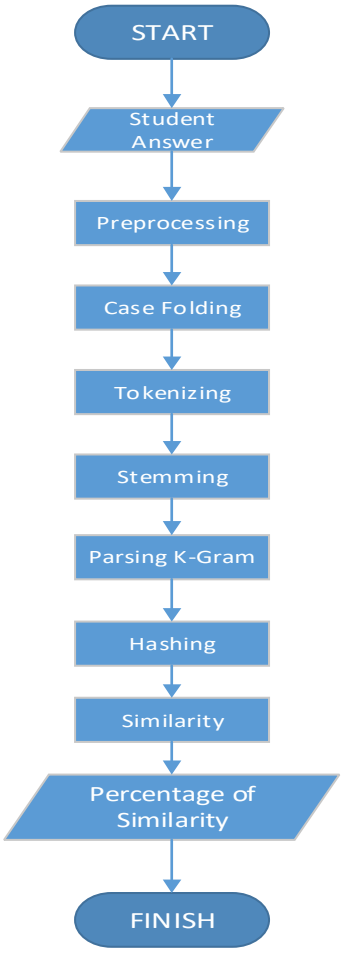

Fig. 1 Automatic essay correction diagram

There are several processes in preprocessing, such as case folding, tokenizing, and stemming. This process aims to take text and remove symbols, special characters, space, punctuation, and change uppercase letters to lowercase letters. In this study, the users are elementary school students, so the preprocessing stage is essential. Students at this level, when writing essay answers, tend not to pay attention to punctuation marks and case letters. Our system automatically changes all answers to lowercase in the folding case process.

The tokenizing stage is cutting the input string based on the word arrangement, then eliminating the delimiter. Our system contributes to automatically removing unnecessary punctuation from student answer sheets. In essay questions, students with a lower-order thinking level tend to ignore the use of punctuation marks. The quality of assessment results can improve with removed the delimiters such as commas, colons, and spaces.

Next, the stemming stage finds basic words from tokenizing results by eliminating word affixes. This stemming process significantly affects the essay assessment results because it selects and compares the keywords that appear with the essay's basic words. At this stage, all affixes that are allowed and not allowed in Indonesian morphological rules are stored in the system's database.

The blended learning system will calculate the similarity value between the student answer and the answer key. Calculations are done with *K*-Gram. *K*-Gram is a series of terms with length *K*. This method takes *K* character continuously in the source text until the end of the document. Furthermore, the hashing step transforms strings into unique values with a fixed length as a string marker. This function value is called a hash function, while the resulting value is called the hash value. If it is not hashed, the character search will be performed on varying character lengths, and there are 26 possibilities for each character.

The system will correct the essay exam value based on the weight of the answers. Ideally, the scoring rubric for the answer key should be developing before administering the essay questions. Student essay exam scores that have been calculated are stored in the database. This blended learning system displays the essay exam scores of students with scores from 0-100. The teacher can see the student's score and the student's answer.

The research population was 20 students at Elementary School in Bengkulu Province. The research was conducted by giving essay questions as many as 12 questions for cycle 1, cycle 2, and cycle 3. The questions were taken from the Heat Transfer Power in Science class for fourth grade in Elementary School.

The test tested with Mean Absolute Error (MAE) and Products Moment Correlation or Pearson Correlation. The test is conducted to evaluate the automated essay system performance. This test using expert assessment results from variables compared with the correction results of the blended learning system.

A. Mean Absolute Error (MAE)

Mean Absolute Error is absolute errors, regardless of positive or negative signs. The original error value is usually not averaged as the error size because there are positive and some negative. So that if the sum of the errors is small, a positive error will be reduced by a negative error [20]. The error needs to be made an absolute number (neglected positive and negative signs) to avoid this. The error is considered small if not exceeding 30%.

$$MAE(\bar{x}) = \frac{\sum_{i=1}^{n} |\boldsymbol{X_i} - \boldsymbol{Y_i}|}{\boldsymbol{N}} \quad (2)$$

Mean Absolute Error formulated with equation (2), where *X* for the first variable (expert correction score), *Y* for the second variable (system correction score), and *N* for data.

B. Products Moment Correlation

Products Moment Correlation or Pearson Correlation is one technique to find a correlation between dependent and independent variables. In this study, the dependent variable is the expert correction score, while the independent variable is the correction score generated from the blended learning system. Products Moment Correlation is a statistical test tool used to test associative hypotheses (Test Relations) of two variables if the data is an interval or ratio scale. This

correlation is obtained by finding the moment's product from the correlated variable. The correlation between two variables can be seen by looking at the correlation index number given the symbol *r*. Correlations are given an index with lowercase letters for variables. If the first variable is given the symbol index *X* for expert correction score and the second variable is given the symbol *Y* for system correction score. The correlation index number is expressed by the symbol r_{xy}. Based on Atoum [21], Products Moment Correlation can be formulated mathematically with equation (3).

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}} \quad (3)$$

Products Moment Correlation or Pearson Correlation is used to explaining the degree between the independent variable and the dependent variable with a value: $-1 \leq r \leq 1$. Guidelines for the correlation coefficient interpretations can be seen in Table I.

- If the value of $r = -1$ or close to -1, then the correlation between the two variables is very strong and negative, which means the relationship between the two variables is in the opposite direction. If the *X* rises, then the *Y* will decrease or vice versa.
- If the value of $r = 0$ or close to 0, then the correlation between the two variables is very weak or no correlation at all.
- If the value of $r = 1$ or close to 1, then the correlation between two variables is very strong and positive, which means the relationship between two variables is in the same direction. If the *X* rises, then the *Y* also rises or vice versa.

TABLE I
THE CORRELATION COEFFICIENT INTERPRETATION

Interval Coefficient	Level
0.00 – 0.19	Very Low
0.20 – 0.39	Low
0.40 – 0.59	Moderate
0.60 – 0.79	Good
0.80 – 1.00	Very Good

III. RESULT AND DISCUSSION

The automated essay correction process determines whether the system can access the answer according to the elementary school teacher as an expert judgment and the system's essay score's accuracy level. The preprocessing of Rabin Karp changes the string into a hash. The first process is done by case-folding (see Fig. 2), tokenizing (see Fig. 3), and stemming (see Fig. 4) in the Indonesian language.

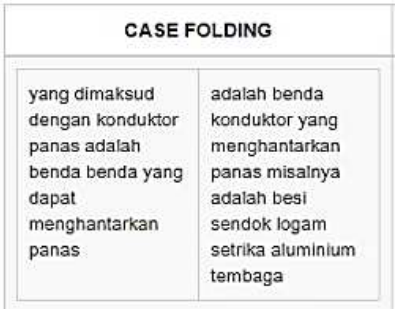

Fig. 2 Case folding process results

Fig. 3 Tokenizing process results

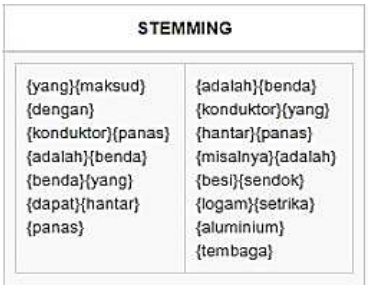

Fig. 4 Stemming process results

The automatic essay correction for elementary school blended learning is presented in the web platform with two privileges: teacher and student. The teacher is responsible for providing questions and answers key. The management exam can be seen in Fig 5.

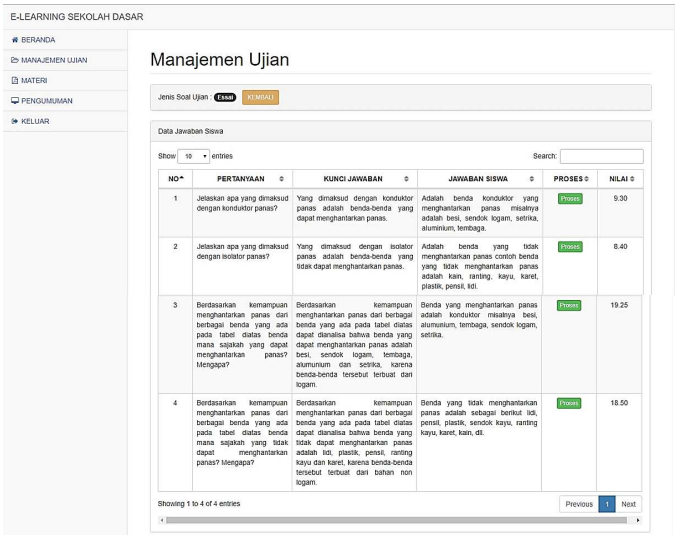

Fig. 5 Exam management and answer key

The statement of each question is presented to students. They must answer within the text editor (Fig. 6). After the answer has been written and the test finished, the system provided a total score to the student.

Fig. 6 Score results of the automatic essay correction when students submit the answer

A. Cycle 1

Table II shows the results of expert and system values for each *k*-gram for 20 students. There are differences between expert and system scores in Cycle 1. The system scores close to the average score of the expert in *k*-gram = 3. The average error and correlation for each *k*-gram are k-gram = 2 with error 10.12 and correlation 0.77, *k*-gram = 3 with error 8.82 and correlation 0.69, *k*-gram = 4 with error 10.64 and correlation 0.63, and *k*-gram = 5 with error 12.76 and correlation 0.69.

TABLE II
CYCLE 1 TESTING

Student	Expert Score	2-Gram	3-Gram	4-Gram	5-Gram
1	75.00	75.38	68.00	62.00	52.88
2	56.00	57.94	48.13	41.19	38.00
3	75.00	73.50	64.56	59.63	53.94
4	68.00	70.13	61.81	52.50	48.38
5	38.00	64.44	56.31	53.69	45.19
6	70.00	69.31	58.06	51.50	48.06
7	66.00	72.06	58.38	52.38	46.44
8	36.00	45.63	36.00	31.38	28.25
9	60.00	69.81	55.44	49.56	42.63
10	71.00	80.81	73.75	70.19	63.94
11	55.00	64.38	56.50	50.31	46.50
12	38.00	64.75	53.94	49.44	41.50
13	70.00	75.56	65.31	60.00	54.44
14	50.00	69.44	61.38	49.69	49.00
15	38.00	56.88	49.50	42.81	37.38
16	39.00	49.31	41.31	36.88	33.31
17	70.00	64.56	51.69	41.44	39.56
18	70.00	68.38	57.31	48.88	47.56
19	62.00	77.06	70.19	63.63	58.63
20	38.00	59.56	51.19	43.81	40.38
Average	57.25	66.44	56.94	50.55	45.80
Highest score	75.00	80.81	73.75	70.19	63.94
Lowest score	36.00	45.63	36.00	31.38	28.25
Average error		10.12	8.82	10.64	12.76
Correlation		0.77	0.69	0.63	0.69

B. Cycle 2

Table III shown the expert and system score for each k-gram from 20 students. There are differences between expert and system scores. It can be seen that the system scores close to the expert score in *k*-gram = 5. The average error and correlation for each *k*-gram are *k*-gram = 2 with error 22.51 and correlation 0.91, *k*-gram = 3 with error 16.11 and correlation 0.85, *k*-gram = 4 error 10.33 and correlation 0.81 and *k*-gram = 5 with error 7.32 and correlation 0.75.

TABLE III
CYCLE 2 TESTING

Student	Expert Score	2-Gram	3-Gram	4-Gram	5-Gram
1	50.00	78.75	72.25	65.50	58.75
2	20.00	54.25	43.00	35.00	34.25
3	60.00	84.75	81.50	76.50	72.00
4	50.00	67.75	59.75	52.75	47.75
5	45.00	70.25	63.50	57.00	52.00
6	58.00	80.25	78.00	71.50	67.00
7	50.00	67.50	62.00	55.50	47.25
8	41.00	69.75	58.00	50.50	45.75
9	60.00	77.25	71.75	67.25	63.75
10	50.00	74.00	65.00	57.75	51.25
11	60.00	72.25	68.50	54.25	45.50
12	60.00	84.25	80.00	73.00	69.00
13	60.00	78.00	72.00	68.00	62.75
14	44.00	52.75	47.25	40.25	36.00
15	60.00	81.25	77.50	70.00	64.75
16	60.00	80.25	74.50	70.25	65.75
17	0.00	0.00	0.00	0.00	0.00
18	42.00	72.25	62.75	56.00	50.75
19	50.00	80.75	75.25	71.00	66.75
20	50.00	71.50	63.50	55.25	47.00
Average	51.05	73.57	67.16	60.38	55.16
Highest score	60.00	84.75	81.50	76.50	72.00
Lowest score	0.00	0.00	0.00	0.00	0.00
Average error		22.51	16.11	10.33	7.32
Correlation		0.91	0.85	0.81	0.75

C. Cycle 3

There are differences between expert and system scores in table IV. The system score is close to the expert class in *k*-gram = 5. The average error and correlation for each *k*-gram are *k*-gram = 2 with error 18.02 and correlation 0.65, *k*-gram = 3 with error 8.43 and correlation 0.80, *k*-gram = 4 with error 5.31 and correlation 0.74 and *k*-gram = 5 with error 5.09 and correlation 0.66.

TABLE IV
CYCLE 3 TESTING

Student	Expert Score	2-Gram	3-Gram	4-Gram	5-Gram
1	65.00	66.63	61.19	54.63	47.50
2	0.00	0.00	0.00	0.00	0.00
3	50.00	62.19	51.88	48.81	40.88
4	43.00	65.88	56.06	50.13	41.56
5	47.00	62.88	52.31	51.19	45.50
6	49.00	62.25	55.38	48.19	41.44
7	27.00	56.00	42.13	39.25	33.00
8	45.00	62.81	54.00	48.75	44.06
9	35.00	55.75	46.38	42.31	38.94
10	46.00	63.38	56.50	53.19	44.50
11	44.00	64.00	52.81	47.88	40.81
12	45.00	60.19	47.06	41.38	34.31
13	42.00	67.69	56.56	50.81	46.88
14	45.00	64.38	53.56	50.56	43.63
15	37.00	63.56	51.13	43.50	36.63
16	50.00	66.13	56.19	53.06	46.06
17	56.00	67.81	60.31	55.75	47.81
18	42.00	60.44	54.13	49.69	44.50
19	43.00	59.88	47.44	40.38	33.88
20	45.00	66.44	53.56	49.63	42.13
Average	45.00	63.07	53.08	48.37	41.79
Highest score	65.00	67.81	61.19	55.75	47.81
Lowest score	0.00	0.00	0.00	0.00	0.00
Average error		18.02	8.43	5.31	5.09
Correlation		0.65	0.80	0.74	0.66

The results with *K*-gram = 5 on the expert and system score in Cycle 3 can be seen in Table V, where *X* is expert assessment results and *Y* is system assessment results.

TABLE V
DIFFERENCES BETWEEN EXPERT AND SYSTEM ASSESSMENT IN CYCLE 3

Student	$[X]$	$[Y]$	$[X-Y$	$[X^2]$	$[Y^2]$	$[XY]$
1	65.00	47.50	17.50	4225.00	2256.25	3087.50
2	0.00	0.00	0.00	0.00	0.00	0.00
3	50.00	40.88	9.12	2500.00	1671.17	2044.00
4	43.00	41.56	1.44	1849.00	1727.23	1787.08
5	47.00	45.50	1.50	2209.00	2070.25	2138.50
6	49.00	41.44	7.56	2401.00	1717.27	2030.56
7	27.00	33.00	6.00	729.00	1089.00	891.00

8	45.00	44.06	0.94	2025.00	1941.28	1982.70
9	35.00	38.94	3.94	1225.00	1516.32	1362.90
10	46.00	44.50	1.50	2116.00	1980.25	2047.00
11	44.00	40.81	3.19	1936.00	1665.46	1795.64
12	45.00	34.31	10.69	2025.00	1177.18	1543.95
13	42.00	46.88	4.88	1764.00	2197.73	1968.96
14	45.00	43.63	1.37	2025.00	1903.58	1963.35
15	37.00	36.63	0.37	1369.00	1341.76	1355.31
16	50.00	46.06	3.94	2500.00	2121.52	2303.00
17	56.00	47.81	8.19	3136.00	2285.80	2677.36
18	42.00	44.50	2.50	1764.00	1980.25	1869.00
19	43.00	33.88	9.12	1849.00	1147.85	1456.84
20	45.00	42.13	2.87	2025.00	1774.94	1895.85
Total	856.00	794.02	96.62	39672.00	33565.10	36200.50

We verify the influences on the average error and the Pearson Correlation. The evaluation of automatic essay correction is needed to assess how the Rabin Karp method works well in scoring the student's answer. This evaluation compares the system score generated automatically from the blended learning system with the expert as a human judgment or lecturer. The result of MAE is the average score difference between system and expert. The correlation coefficient interpretation (r) showed a score of 0.66 at a good level (Table I).

$$MAE(\bar{x}) = \frac{\sum_{i=1}^{n}|X_i - Y_i|}{N} = \frac{96.62}{19} = 5.09$$

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}}$$
$$= \frac{(19\times 36200.50) - ((856)\times(794.02))}{\sqrt{\{(19\times 39672) - (856)^2\}\{(19\times 33565.10) - (794.02)^2\}}}$$
$$= \frac{8128.38}{12364.65} = 0.66$$

Based on system performance testing from three cycles that have been carried out that 5-gram produces the system assessment closest to the expert assessment results. The cause of small accuracy in the experiment with 2-gram and 3-gram was the frequencies of term words in the student's answer key. The lowest difference between the Rabin Karp algorithm and the teacher value is 0.66 in Cycle 3 and 5-gram (Table VI).

TABLE VI
SYSTEM PERFORMANCE RESULTS (*K*-GRAM = 5)

Cycle	Average Error	Correlation
1	12.76	0.69
2	7.32	0.75
3	5.09	0.66
Average	8.39	0.70

Table 6 shows the average errors and correlations from the system performance assessment using 5-gram. The average error is relatively small, about 8.39 with a correlation of 0.70, which means having a good correlation or relationship between expert and system assessment. Several factors influence the Rabin-Karp algorithm performance, such as the amount of content, stemming and preprocessing time, and the value of *k*-gram.

According to our observations, blended learning for elementary students can motivate students to participate more actively. Students were more productive to written essays and written scientific answers. They not only develop communication skills but also writing skills.

This essay assessment program is flexible for evaluating various interrelated question content, such as Science, Technology, Engineering, Art, and Mathematics (STEAM). Student's essays can describe science as knowledge through technology, analyze the design in engineering, and explain the arithmetic operations on the mathematics aspect. Further research on this system can be applied to accommodate student learning outcomes integrated with STEAM elements through blended learning.

Finally, it was observed that automatic essay correction influences the students learning. They are paid more attention to evaluation with learning media in blended learning because it displays learning information more interactive [3], [22].

IV. CONCLUSION

The application of essay correction is still challenging to enhance reliable and valid markings. This paper has presented an automatic essay correction tool that assesses student's descriptive answers in blended learning for elementary school. The essay score is generated automatically depending on the Rabin Karp results for each correct keyword usage.

Based on experiments, too small or too large *K*-grams can make the similarity value inaccurate, and the processing time becomes too long or too fast. The preprocessing stage affects the accuracy of similarity and processing time. This process produces a similarity value that tends to be more accurate than without preprocessing. The correction system can provide scores closest to expert scores using grams to find matches of the same word or phrase from the student's answer text. The Rabin-Karp method has been able to examine essay questions automatically with the answer key to the database. The difference score between the automatic system and the expert is very close. This system has a better score distribution, so not significantly different from the teacher or expert assessment.

The blended learning model can shift the learning principle dynamically from teacher-center learning to student center learning. The use of automatic essay correction in blended learning can improve the writing skills of elementary school students. This assessment can evaluate and measure student learning outcomes with integrated subject such as Science, Technology, Engineering, Art, and Mathematics (STEAM). It also contributes to increasing motivation and self-development for students in the digital educational environment in the 4.0 era.

ACKNOWLEDGMENT

We are grateful to the Directorate of Higher Education through the Ministry of Research, Technology, and Higher Education, Republic Indonesia, for financial support in this Fundamental Research scheme.

REFERENCES

[1] H. M. Balaha and M. M. Saafan, "Automatic exam correction framework (AECF) for the MCQS, essays, and equations matching," *IEEE Access*, vol. 9, no. 1, pp. 32368–32389, 2021, doi: 10.1109/ACCESS.2021.3060940.

[2] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "E-Learning: Challenges and Research Opportunities Using Machine Learning Data Analytics," *IEEE Access*, vol. 6, pp. 39117–39138, 2018, doi: 10.1109/ACCESS.2018.2851790.

[3] C. Senturk, "Effects of the blended learning model on preservice teachers' academic achievements and twenty-first century skills," *Educ. Inf. Technol.*, vol. 26, no. 1, pp. 35–48, 2021, doi: 10.1007/s10639-020-10340-y.

[4] H. Syam, M. Basri, A. Abduh, A. A. Patak, and Rosmaladewi, "Hybrid e-learning in Industrial Revolution 4.0 for Indonesia higher education," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 9, no. 4, pp. 1183–1189, 2019, doi: 10.18517/ijaseit.9.4.9411.

[5] M. Jaya Saragih, R. Mas Rizky Yohannes Cristanto, Y. Effendi, and E. M. Zamzami, "Application of Blended Learning Supporting Digital Education 4.0," *J. Phys. Conf. Ser.*, vol. 1566, no. 1, pp. 0–6, 2020, doi: 10.1088/1742-6596/1566/1/012044.

[6] J. Moreno and A. F. Pineda, "A Framework for Automated Formative Assessment in Mathematics Courses," *IEEE Access*, vol. 8, pp. 30152–30159, 2020, doi: 10.1109/ACCESS.2020.2973026.

[7] M. Liu, Y. Li, W. Xu, and L. Liu, "Automated Essay Feedback Generation and Its Impact on Revision," *IEEE Trans. Learn. Technol.*, vol. 10, no. 4, pp. 502–513, 2017.

[8] N. T. Thomas, A. Kumar, and K. Bijlani, "Automatic Answer Assessment in LMS Using Latent Semantic Analysis," *Procedia Comput. Sci.*, vol. 58, pp. 257–264, 2015, doi: 10.1016/j.procs.2015.08.019.

[9] G. Liang, B. W. On, D. Jeong, H. C. Kim, and G. S. Choi, "Automated essay scoring: A siamese bidirectional LSTM neural network architecture," *Symmetry (Basel).*, vol. 10, no. 12, pp. 1–16, 2018, doi: 10.3390/sym10120682.

[10] R. Setiadi Citawan, V. Christanti Mawardi, and B. Mulyawan, "Automatic Essay Scoring in E-learning System Using LSA Method with N-Gram Feature for Bahasa Indonesia," in *MATEC Web of Conferences*, 2018, vol. 164, doi: 10.1051/matecconf/201816401037.

[11] A. Yudhana, S. -, and A. Djalil, "Implementation of Pattern Matching Algorithm for Portable Document Format," in *International Journal of Advanced Computer Science and Applications*, 2017, vol. 8, no. 11, pp. 509–512, doi: 10.14569/ijacsa.2017.081162.

[12] C. L. Frinhani, S. A. A. De Freitas, M. V. Fernandes, and E. Dias Canedo, "An automatic essay correction for an active learning environment," *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA*, vol. 0, pp. 1–6, 2016, doi: 10.1109/AICCSA.2016.7945769.

[13] R. Setiawan, W. Budiharto, I. H. Kartowisastro, and H. Prabowo, "Finding model through latent semantic approach to reveal the topic of discussion in discussion forum," *Educ. Inf. Technol.*, vol. 25, no. 1, pp. 31–50, 2020, doi: 10.1007/s10639-019-09901-7.

[14] K. Kim, Y. J. Choi, M. Kim, J. W. Lee, D. S. Park, and N. Moon, "Teaching-learning activity modeling based on data analysis," *Symmetry (Basel).*, vol. 7, no. 1, pp. 206–219, 2015, doi: 10.3390/sym7010206.

[15] A. Petchprasert, "Utilizing an automated tool analysis to evaluate EFL students' writing performances," *Asian-Pacific J. Second Foreign Lang. Educ.*, vol. 6, no. 1, pp. 1–16, 2021, doi: 10.1186/s40862-020-00107-w.

[16] J. M. Carman, "Blended Learning Design: Five Key Ingredients," in *Proceedings of the Seventh IASTED International Conference on Computers and Advanced Technology in Education*, 2002, no. October, pp. 491–496.

[17] R. M. Karp and M. O. Rabin, "Efficient randomized pattern-matching algorithms," *IBM J. Res. Dev.*, vol. 31, no. 2, pp. 249–260, 1987.

[18] S. I. Hakak, A. Kamsin, P. Shivakumara, G. A. Gilkar, W. Z. Khan, and M. Imran, "Exact String Matching Algorithms: Survey, Issues, and Future Research Directions," *IEEE Access*, vol. 7, pp. 69614–69637, 2019, doi: 10.1109/ACCESS.2019.2914071.

[19] V. Verma and R. K. Aggarwal, "A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, pp. 1–16, 2020, doi: 10.1007/s13278-020-00660-9.

[20] W. Wang and Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 324, no. 1, 2018, doi: 10.1088/1757-899X/324/1/012049.

[21] I. Atoum, "Scaled Pearson's Correlation Coefficient for Evaluating Text Similarity Measures," *Mod. Appl. Sci.*, vol. 13, no. 10, p. 26, 2019, doi: 10.5539/mas.v13n10p26.

[22] E. W. Winarni and E. P. Purwandari, "The effectiveness of turtle mobile learning application for scientific literacy in elementary school," *J. Educ. e-Learning Res.*, vol. 6, no. 4, pp. 156–161, 2019, doi: 10.20448/journal.509.2019.64.156.161.