

Using Conventional Rubric and Coh-Metrix to Assess EFL Students' Essays

Monica Kristin Mahadini

Sebelas Maret University
Email: monicakristin12@gmail.com

Endang Setyaningsih

Sebelas Maret University
Email: endang_setyaningsih@staff.uns.ac.id

Teguh Sarosa

Sebelas Maret University
Email: teguhsrs@staff.uns.ac.id

Received: 02 February 2021

Reviewed: 1 September 2021-29 November 2021

Accepted: 7 December 2021

Abstract

The conventional scoring rubric has been used as a tool to assess students' writing for years and recently there is growing attention on the use of automated assessment tools. While studies highlighting the use of the tools per se are numerous, not much is available on the use of the combined tools. This study addresses the gap by examining 20 university EFL students' essays using both assessment tools. The essays were examined based on a conventional rubric which contains five aspects and two categories covering eleven indices of an automated tool called Coh-Metrix. The results of the examination were interpreted and information generated from the tools are compared. The study reveals that the use of both conventional and automated tools laid a more comprehensive picture of the students' essays. The results of the assessment are useful to inform students and teachers on areas that need attention in the writing instruction. Gain and loss on the use of the tools are explained.

Keywords: Coh-Metrix; EFL; Essay; Rubric; Writing

Introduction

Students' writing is conventionally assessed using a pre-developed rubric that usually covers five aspects, i.e., grammar, vocabulary, mechanics, fluency, and form/organization (Hughes, 1989). The previous studies in the use of the conventional rubric reported three points of benefits. First, for the teacher, it can be used to identify the strengths and weaknesses of every aspect of students' writing. The use of rubrics makes it possible for teachers to keep complete detailed records of each student's progress. Detailed feedback based on rubrics is useful for accurately analyzing where students' strengths and weaknesses are located (Abbas, 2017; Gulzar et al., 2017; Moskal, 2000). Second, for the students, detailed feedback helps them improve or revise their writing (Gulzar et al., 2017). Third, the rubric fences personal subjective opinion in

assessing students' writing and keeps the score transparent and fair, and develops a smooth learning path (Gulzar et al., 2017; Moskal, 2000).

Today, the alternative to conventional assessment appears in the form of automated assessment, and one of them is through Coh-Metrix. Coh-Metrix is a tool that provides a variety of computational linguistic indices to meet the evolving needs of comprehensive and automated text analysis (Graesser et al., 2011). Coh-Metrix is an online platform that can be used to analyze text seen from the linguistic elements (Graesser et al., 2004). Coh-Metrix 3.0 provides 110 indices which are classified into 11 groups. Assessment of students' writing using Coh-Metrix has been conducted in different contexts. Lei et al. (2014) used Coh-Metrix to analyze students' essay texts in the General Core Technology curriculum course. Zedelius et al. (2018) used Coh Metrix to evaluate college students' creative writing and Tortorelli (2019) used the tool to analyze the text complexity of grade 2 elementary school students. The results of the Coh-Metrix analysis were commonly used to evaluate and describe the need to improve students' writing skills.

While previous studies have showcased the use of the conventional rubric and Coh-Metrix per se to examine different types of text in different contexts, the attempt to combine the use of the two tools to assess students' writing is absent. This present study aims to shed some light on the potential of combining the use of rubric and Coh-Metrix to provide a better analytical assessment of the students' writings. This study examines 20 essays written by university EFL students in Indonesia. The essays were examined based on a conventional rubric which contains five aspects and two categories covering eleven indices of Coh-Metrix. The results of the analysis are interpreted to evaluate students' ability to write an essay. It highlights the information of the assessment that can potentially be used to help students improve their writing and inform the teacher on areas that need attention in the writing instruction.

Literature review

Assessing essay text

Assessment of an essay text has been an area of interest for many educational researchers. Traditionally in assessing students' ability to write essay texts, the teacher evaluates manually by grading based on the scoring rubric that has been available from several experts. There are indicators by some experts that usually used to evaluate and assess students' text such as Brown (2004: 246) states five aspects of writing as follows: content, organization, vocabulary, syntax, and mechanics. Johnson and Johnson (2002: 75) state that to evaluate a composition, the components are: organization, cohesion, vocabulary, sentence structure, and spelling and punctuation. Hughes (1996: 91) mentions five aspects of writing: grammar, vocabulary, mechanics, fluency, and form (organization). Kitao (1996: 2) mentions that the ability to write involves the components of writing as follows: grammatical ability, lexical abilities, mechanical, stylistic skills, organizational skills, and judgment appropriateness. Nunan (1998: 36) states that writing is an extremely complex cognitive activity in which the writer is required to demonstrate control of number variables, which include: control of content, format, sentence structure, vocabulary, punctuation, spelling, and letter formation. It can be concluded that the major indicators commonly used to test and score the students' writing skills include: content, organization of ideas, grammar, vocabulary, and mechanics. A more detailed description of the indicators offered by Jacob (1981), and has been relatively popular as well as widely used writing teachers. (See Appendix)

The main issue with conventional writing assessment using a rubric is the potential scorer's subjectivity. To address this challenge, it suggested having an inter-rater (two or more scorers).

Another way that is also commonly mentioned by experts is to score the students' writing twice at different times by the same scorer (intra-rater). The most ideal way is combining the two methods: score the writing by intra-rater and inter-rater.

Coh-Metrix Analysis

In recent years, the use of an automatic grading system for analyzing written texts, including responses generated by students is growing. (Magliano and Graesser 2012; Shermis et al. 2016). Specifically, automatic assessment has been used to assess students' writing abilities that vary in age and language skills (Crossley et al. 2016; Weigle 2013; Wilson et al. 2016). An automatic assessment has the potential to contribute to understanding the teaching needs of poor writers because of its accuracy and consistency in analyzing the linguistic and structural aspects of writing (Deane and Quinlan 2010).

One of the prominent automatic scoring tools that are widely used is a computational linguistic tool called Coh-Metrix. It was and is continuously developed by McNamara, Graesser, McCarthy, and Cai (2004). Coh-Metrix is a tool that provides a wide range of computational linguistic indexes to meet the growing need for comprehensive and automatic text analyses. Coh-Metrix uses lexicons, a syntactic parser, LSA, and several other components that are widely used in computational linguistics (Graesser et al., 2004). Coh-Metrix is a tool that provides a variety of computational linguistic indices to meet the evolving needs of comprehensive and automated text analysis. Coh-Metrix is a tool that provides a variety of computational linguistic indices to meet the evolving needs of comprehensive and automated text analysis (Graesser et al., 2011).

Coh-Metrix has several tools to analyze various texts. But in this study, we used the Coh-Metrix 3.0 available at <http://141.225.41.245/CohMetrix2017/> and Coh-Metrix Text Easability Assessment at <http://tea.cohmetrix.com> to evaluate the results of students' essay texts. Eleven categories and 110 indices can be used to assess in this tool. The categories are descriptives, text easability principal components scores, referential cohesion, latent semantic analysis, lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density, word information, and Readability (Dowell et al., 2016; Quispesaravia et al., 2016).

To limit the scope of this study, we used two categories and nine indices to assess the students' essays. The first category is 'descriptive' that includes four indices: number of paragraphs, number of sentences, number of words, and number of sentences in a paragraph. The second category is the "Text Easability Principles Component" score that includes five indices: narrativity, syntactic simplicity, word concreteness, referential, and deep cohesion. We chose these indices from Coh-Metrix because they can be conveniently linked with four out of five aspects in the conventional scoring rubric i.e. 'content', 'vocabulary', 'grammar', and 'organization'.

Evaluating student texts through computerization has several advantages. Due to the automation, various types and levels of the students' texts can be evaluated automatically within a short time. However, Coh-Metrix has limits note that both tools are limited to about 15,000 characters (Dowell et al., 2016). Besides that, evaluating text using Coh-Metrix also requires an interpretation that must be done by the teacher between the type of text being evaluated and the aspects that exist in Coh-Metrix. For example, a low aspect of narrativity in the argumentative text is needed because the argumentative text is not story-like text, so narrativity must be low in aspect.

Research method

This research employs the design of document analysis. Document analysis is a form of qualitative method in which documents are interpreted to give voice and meaning around an assessment topic (Brown, 2009). It covers five steps namely identifying and collecting data (20

students' essays), determining coding categories, coding the content, checking validity and reliability, and analyzing and presenting results. These steps are conducted manually when data were examined using scoring rubric. But when data were input to the Coh-Metrix, the steps were conducted by the system automatically. Implementing this design can reveal the quality of the students' essay as well as the strengths and weaknesses of the chosen writing assessment tools, namely scoring rubric and Coh-Metrix, which will be beneficial for writing teachers.

A total of 81 EFL students' essays were collected for this study and 20 essays were taken randomly as the samples for further analysis using both conventional scoring rubric and automated Coh-Metrix.

To minimize subjectivity and to enhance the validity of the measurement and data, inter-rater was involved to assess the students' writing using the conventional scoring rubric. The inter-rater was selected based on certain criteria i.e. English teacher who has sufficient experience in teaching and is familiar with the use of a scoring rubric. To ensure the inter-rater reliability, the Kappa coefficient calculation was carried out using SPSS 26.

Symmetric Measures					
		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.772	.115	5.119	.000
N of Valid Cases		20			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

Figure 1. The output of Kappa coefficient computation (SPSS ver. 26)

Based on Landis and Koch (1977), Kappa value of more than 0.61 is indicated as substantial reliability. Thus, it can be interpreted that the inter-rater reliability of this study is substantial. The rubric used in this study was developed/ adapted/ adopted from and it highlights five aspects namely Content, Mechanic, Vocabulary, Grammar, and Organization. The assessment using the rubric was carried out concurrently with the assessment using Coh Metrix.

The assessment of the students' essays using Coh Metrix was carried out by inputting the 20 essays into the system one by one. To get a quantitative description of a text, the file of the text was copied and pasted on to a slot provided at <http://141.225.41.245/CohMetrix2017/>. Once all texts have been measured, the result of the selected indices i.e. number of words, number of sentences, and number of paragraphs were exported into excel. Meanwhile to get the measure of the five indices of Text Easibility Assessment (TEA) i.e. narrativity, syntactic simplicity, word concreteness, referential, and deep cohesion, the text was copied and pasted onto a slot provided at <http://tea.cohmetrix.com>. The generated data were then exported to a separate excel. The reliability of the Coh-Metrix assessment is assumed to have been built-in in the system.

The data analysis begins with the tabulation of all data generated from the Coh-Metrix (both descriptive and TEA) and data from raters. The data are then computed to get descriptive statistics (average/ mean). The next step is looking for patterns and peculiarity. The datum that is above or below average is carefully examined. The emerging general patterns of the data were discussed with the existing theory and/ or previous findings. At the same time, the students' texts were closely analyzed and marked, mainly manually with the aid of the rubric. Errors were

identified, counted, and clustered. Parts of the texts that correspond to the result of the scoring rubric and Coh-Matrix were also marked and organized to get patterns or general themes. For example, grammar errors were identified, given symbol or code, then categorized into clusters of similar errors; also, conjunctions and pronoun and overlap technique that were found in the texts are marked and clustered into evidence for the cohesion aspect in the TEA.

Result and discussion

Assessment of student's essays using the conventional scoring rubric

Five aspects of writing are included in the conventional scoring rubric that was used in this study: content, mechanic, vocabulary, grammar, and organization. Within each aspect, there are four categories of assessment: Excellent for 4, Good for 3, Fair for 2, and Poor for 1. These aspects are found in the scoring rubric which is adopted by Jacob, et al. (1981: 90) in (Ngadiso, 2013).

Table 1. Frequency of scoring rubric aspect.

Category	Content	Mechanic	Vocabulary	Grammar	Organization of idea
Excellent	8 (40%)	4 (20%)	2 (10%)	2 (10%)	9 (45%)
Good	12 (60%)	13 (65%)	17 (85%)	14 (70%)	11 (55%)
Fair	0	3 (15%)	1 (5%)	4 (20%)	0
Poor	0	0	0	0	0

Table 1 displays the students are relatively good at the area of content and organization but they are relatively struggling in the mechanic, grammar, and vocabulary. The analysis found that the students' essays are relevant to the topic but most of them are not well developed. Essays labeled as 'excellent' have at least five sentences in each paragraph that support the main idea but essays labeled as 'good' were lacking supporting details. The organization of the ideas was also relatively good. Careful examination of the students' writing reveals that the ideas of the essays are organized, logically, and coherent. Besides, the writings are presented in the appropriate staging of an argumentative essay: introduction, main points, and conclusion. However, more than half of the essays still have several issues in terms of the organization such as repeated statements, loose organization but having standout main idea, fairly limited idea development, and having logical but incomplete ordering.

Three issues that are found in the students' essays are the mechanic, grammar, and vocabulary; three aspects that can be closely linked to problems commonly faced by EFL students. The result of the analysis bares that the students often made errors in spelling, for example 'through out' (Text 7), 'distrupt' (Text 3), and 'anually' (Text 17); punctuation, for example, '...that stands in an area, because it will have a positive impact...(Text 11); capitalization, for example '...preparation of the law. right system to be implemented...' (Text 3) and *The myth of Pamoksan Prabu Brawijaya* (Text 1). In the grammar aspect, identified problems include, among others, error in agreement, missing article, incorrect use of the pronoun, and error in the construction of simple/complex sentences (e.g. lacking verb or subject). Here are some examples of sentences with grammar problem:

1. *The results exported to China and Thailand* (Text 11; lacking verb)
2. *Contains high antioxidant that...* (Text 9; lacking subject)
3. *...and has carried out continuous improvement of irrigation canals* (Text 11; missing article);

4. . . ., this plan really need to be realized (Text 13; agreement)

Meanwhile, issues found in the area of ‘vocabulary’ are lack of vocabulary and incorrect word choice and usage. Some concept like *begah* and specific name of herbs such as *Temulawak* and *Brotowali* have no English equivalent and usually needs extended explanation. However, some students were unable to provide sufficient descriptions or find the closest equivalent. There are also cases in which the students used highly technical terms such as *diarylheptanoid* and *tautomeric compound* without ample understanding of the concept. As a result, some of the words are used in an inappropriate context; making the sentence in which the words are used becomes confusing and obscure. Meanwhile, samples of incorrect use of word choice and usage are observable in the following examples. “...was happen, like regent or president (instead of regional or presidential) election (Text 5) and “Not only increasing the economy (instead of economic) status” (Text 13)

The results of the examination of the EFL students’ essays are similar to those found by Abbas (2017) who also examines university EFL students’ essays in Indonesia using a rubric. In his study, Abbas (2017) reported that the EFL students have problems in the aspect of content (generating ideas for thesis statement) and mechanics (spelling, punctuation, capitalization), and organization (paragraphing). This implies that the use of the scoring rubric has been able to provide both teachers and students with information on aspects of writing that needs improvement. Having identified difficulties faced by each student, the teacher can make further observations to each student who is deemed insufficient and plan an action plan or writing strategy to help students improve their writing. (Abbas, 2017; Gulzar et al., 2017). More, Gulzar et al. (2017) noted that using a rubric allows the teacher to keep detailed notes on the progress of each student. The detailed feedback on the rubric is useful for analyzing the strengths and weaknesses of the students accurately. Rubric fences personal subjective opinion in assessing student writing and keeps the score transparent and fair and develops a smooth learning path (Gulzar et al., 2017; Moskal, 2000). Lastly, it has a big impact on teaching as well as it helps teachers to evaluate their teaching performance after tracking student performance.

Regardless of the gain, the use of the scoring rubric in assessing students' texts also has several losses. The teacher takes a long time to assess the students' texts. Teachers must check and mark papers while following a specific rubric on each student's work. When this was done on a large scale, it would take a lot of time. Besides, there is subjectivity that can occur when assessing. To address this challenge, it suggested having an inter-rater (two or more scorers). Another way that is also commonly mentioned by experts is to score the students' writing twice at different times by the same scorer (intra-rater) (Ngadiso, 2013). Either way, the time required to do the assessment is usually lengthy and a certain extent of subjectivity remains.

Assessment of students’ essays using Coh-Metrix

There are two categories and nine indices used in assessing students' essay texts. In the descriptive category, we limit the study to four indices accessed using Coh-Metrix 3.0 at <http://141.225.41.245/CohMetrix2017/>. The categories selected are the number of paragraphs, number of sentences, number of words, and number of sentences in a paragraph. Whereas in the Text Easability Principles Component category using the Coh-Metrix Text Easability Assessment at <http://tea.cohmetrix.com> which has five aspects, namely narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion.

Descriptive category

The Coh-Metrix 3.0 provides an automatic descriptive measurement of the texts. Having the text input into the system, data on the quantitative descriptive category can be generated within seconds.

Table 2. Coh-Metrix 3.0 output

No.	Text	Number of Paragraphs	Number of Sentences	Number of Words	Number of Sentences in a Paragraph
1	Text 1	6	68	1,114	11.333
2	Text 2	6	33	544	5.500
3	Text 3	6	57	1,232	9.500
4	Text 4	6	42	848	7.000
5	Text 5	6	45	581	7.500
6	Text 6	6	41	765	6.833
7	Text 7	6	33	710	5.500
8	Text 8	6	42	850	7.000
9	Text 9	6	42	704	7.000
10	Text 10	6	34	818	5.667
11	Text 11	6	55	1,051	9.167
12	Text 12	6	55	964	9.167
13	Text 13	5	21	582	4.200
14	Text 14	6	40	667	6.667
15	Text 15	6	20	600	3.333
16	Text 16	6	50	967	8.333
17	Text 17	6	26	642	4.333
18	Text 18	6	30	837	5.000
19	Text 19	6	38	775	6.333
20	Text 20	6	23	399	3.833
	Average	6	40	783	6.7

This quantitative data is useful for checking whether the students' work has met specific requirements set by the instructor such as the minimum number of paragraphs or length of the essays. For example, Text 13, can be automatically identified as a text that does not meet the requirement that has been set in terms of the number of the paragraph. This information can be used to detect incomplete part of the argumentative essay. Text 13, in this case, fails to provide the minimum number of claims and evidence to build the argument.

Data from the descriptive output of Coh Metrix can also be used to check students' ability to develop ideas into their writing (Lei et al., 2014). In this study, the average number of sentences that the students can write is 39-40 sentences in one text and approximately 6-7 sentences in a paragraph. The average number of words is 783. There are nine students, however, who wrote below the average. These students can be predicted to have issues related to the development of ideas. Nevertheless, this prediction needs to be taken with caution. Students who write a few sentences can produce more words than students who write many sentences. In this case, these students are likely to write complex sentences in which ideas can be well elaborated within smaller number of sentences.

Text easability assessment category

Five aspects of Coh-Metrix Text Easability Assessment (TEA) at <http://tea.cohmetrix.com> are used to analyze the text easability. They are narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion (Graesser et al., 2011). Once the text is uploaded into the system, the Coh Metrix TEA will display the measurement of the five aspects in form of a bar chart. The recap of the TEA output is presented in Table 3.

Table 3. Coh-Metrix TEA output

Text	Narrativity	Syntactic Simplicity	Word Concreteness	Referential Cohesion	Deep Cohesion
Text 1	30%	71%	10%	29%	79%
Text 2	34%	66%	55%	67%	85%
Text 3	22%	75%	9%	18%	87%
Text 4	5%	59%	16%	62%	65%
Text 5	42%	89%	2%	60%	87%
Text 6	27%	61%	9%	65%	83%
Text 7	34%	42%	45%	77%	96%
Text 8	12%	39%	14%	88%	61%
Text 9	6%	59%	45%	22%	97%
Text 10	12%	27%	59%	96%	71%
Text 11	11%	49%	44%	55%	82%
Text 12	14%	49%	14%	68%	70%
Text 13	26%	26%	36%	57%	99%
Text 14	30%	67%	39%	57%	99%
Text 15	20%	25%	63%	82%	69%
Text 16	42%	42%	38%	98%	97%
Text 17	27%	17%	67%	31%	100%
Text 18	38%	21%	13%	59%	91%
Text 19	20%	45%	12%	48%	56%
Text 20	21%	55%	69%	65%	88%
Mean	24%	49%	33%	60%	83%

Table 3 shows that, on average, students' essays are having relatively low narrativity and word concreteness but relatively high syntactic simplicity, referential cohesion, and deep cohesion. Similar to the descriptive output, the TEA output provides quantitative measure and brief description but they are subject to further interpretation. The low narrativity in the output is described as a factor that may make the text difficult to comprehend. However, since the texts being examined are argumentative and not story-like, the low score on the narrativity indices is favored. In non-story-like texts, there should be lots of verbs that convey actions, thoughts, and feelings; all of which make the text more accessible. Apart from that, the text also contains many more pronouns and is sprinkled freely throughout the text. Pronouns personalize as well as reach out and help bring the reader into the text whether it is story-like or not. The low narrativity in the students' argumentative essays may also be caused by the use of technical and non-frequently used words. In this study, technical words related to environmental problems were frequently used to respond to the assigned topic. In this case, the reader's prior domain knowledge should be specifically considered as well. Readers of argumentative essays with a specific topic on an environmental issue are somewhat more specific than general and they should have possessed a certain extent of background knowledge on the issue to help them comprehend the text. The low

narrativity corresponds to the low concreteness score which means more abstract words are found in the essay. More abstract words can mean more abstract ideas, as well (Graesser et al., 2014; McNamara et al., 2014) which suit the genre assigned.

In contrast to the low score of narrativity and word concreteness, the score of syntactic simplicity is high. Text with fewer clauses, words per sentence, and the word before the main verb will give the text a higher score for syntactic simplicity. This means the essays with more simple sentences tend to get higher scores. Close examination of the essays confirms that the majority of ideas in the essays are presented in a simple construction. However, at the college level, students are expected to write more complex sentences. Therefore, the favored obtained score, in this particular case, should be low instead of high.

The Coh-Metrix TEA measurement also shows that the students' essays have high referential and deep cohesion. Based on Graesser et al., (2014) and McNamara et al., (2014), a text with high referential cohesion is indicated with the presence of words and ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader. Meanwhile, a text having high deep cohesion is indicated by the presence of conjunctions that connects ideas or parts within the text. Both referential and deep cohesion serve as tools that may help readers to scaffold comprehension, particularly if the content is challenging.

Insofar, the Coh-Metrix benefits teachers with fast measurement. In several simple steps and relatively short time, teachers can obtain data on students' writing descriptions and easibility. Nevertheless, the result of the Coh-Metrix measurement needs to be interpreted according to the genre, level, and topic of the text being assessed. For example, in this study assessing argumentative texts must be interpreted with several aspects of the Coh-Metrix such as narrativity related to genre, syntactic simplicity related to student level, and word concreteness related to the topic being written. Also, there is a maximum limit of up to 15,000 characters in a single text (Dowell et al., 2016). So the Coh-Metrix is not suitable for long or large-scale texts.

Relationship between the conventional scoring rubric and the Coh-Metrix

Both the conventional scoring rubric and the automated tool, the Coh-Metrix, can be used to assess text. Both tools have relationships found on several indices. Descriptive indices in the Coh-Metrix are related to the content aspect of the scoring rubric. The Coh-Metrix TEA indices are linkable to the grammar, vocabulary, and organization of ideas on the scoring rubric.

Descriptive categories help assess students' text in quantity. The number of sentences, number of words, and number of sentences in a paragraph in a text serve as indicators of whether students can develop their ideas. The content aspect of the scoring rubric needs to be used to ensure whether these students can develop their ideas following the topic given. On the other hand, the descriptive category in the Coh-Metrix also helps teachers to predict if the students' writings meet the expectation. For example, if the students produce a minimum number of words or sentences, or paragraphs, they might be struggling to develop ideas in their writing. In the TEA, word concreteness, syntactic simplicity, deep and referential cohesion are, to a certain extent, provide cues on the vocabulary, the types of sentences (whether simple or complex), and the organization of the texts. The cues may not be detailed but they could provide the necessary early hints on the quality of the texts for further and more careful examination.

In line with Matthews & Wijeyewardene (2018), automatic and human evaluations produce measurements grouped into broad categories of text cohesion, lexical characteristics, and syntactic complexity. One of these categories can be related to the present study, text cohesion. In the conventional scoring rubric, the level of cohesion in a text can be seen and measured in the

organization of the idea category. Whereas in the Coh-Metrix, the cohesion can be seen in the referential score and deep cohesion in the text accessibility principle components score category. Based on these studies, it was stated that the correlation coefficients between the two were positive. If a text has a high percentage of referential cohesion and deep cohesion, it can be concluded that events, ideas, and information from the whole text are tied together. It can be measured by conjunctions and types of words that connect various parts of the text. This percentage can be a reference in scoring the organization of ideas in the scoring rubric.

Conclusion

From the research findings, it can be concluded that the students' writing can be assessed using either the conventional scoring rubric or Coh-Metrix. The result of the examination of the text using the conventional scoring rubric indicates that the students could develop and organize the ideas well but needs further training on the aspect of vocabulary, mechanics, and grammar. The result of the Coh-Metrix analysis shows that the students' essays get low scores in terms of narrativity and word concreteness but get high scores on syntactic simplicity and cohesion (deep and referential). The results of the Coh-Metrix indices score, however, are highly subject to further interpretation. High and low scores do not translate as good and bad. The interpretation depends on several factors such as level and genre.

More importantly, the study also reveals that the use of both the conventional scoring rubric and Coh-Metrix provide a more comprehensive understanding of student essays because each measure has its gain and loss. Measurement of student essays using the conventional scoring rubric is observably time-consuming compared to the Coh-Metrix. In addition, the Coh-Metrix provides more concrete illustration of items that are qualitative on the scoring rubric. In the scoring rubric assessment, a certain degree of subjectivity remains regardless of the presence of inter-rater or the implementation of intra-rater. One reason is the benchmark which is in the form of an explanation in words, while in the Coh-Metrix the results are concrete numeric data. While gaining some advantages in terms of time efficiency and quantification of information, the Coh-Metrix losses the ease of interpretation offered by the scoring rubric. To use Coh-Metrix, data has to be interpreted to become information and the process of interpretation demands critical examination based on several aspects such as the genre, the level, and topic.

These findings imply that while the use of Coh-Metrix to analyze students' writing is to a certain extent more complicated than the conventional scoring rubric because it requires teachers' ability to interpret the data. When deciding to use Coh-Metrix, it is suggested that teachers select indices that they think are most relevant to the current assignment. Since this present study is conducted in a limited context, future investigations on the use of both tools are wide open. In particular, studies that relate the indices of Coh-Metrix and conventional rubric will be very helpful for teachers who wish to combine the tools.

Declaration of conflicting interest

The authors declared no potential conflicts of interest in this work.

Funding acknowledgement

This research received no external funding.

References

Abbas, M. F. F. (2017). "Assessing and evaluating EFL learners' ability in writing academic

- essay". Proceedings of the Fifth International Seminar on English Language and Teaching (ISELT-5). Vol. 5, 257–261.
- Brown, H. D. (2004). *Language assessment principles and classroom practice*. New York: Pearson Education.
- Dowell, N. M. M., Graesser, A. C., & Cai, Z. (2016). "Language and discourse analysis with Coh-Metrix: applications from educational material to learning environments at scale". *Journal of Learning Analytics*, 3(3), 72–95. <https://doi.org/10.18608/jla.2016.33.5>
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). "Coh-Metrix measures text characteristics at multiple levels of language and discourse". *Elementary School Journal*, 115(2), 211–229. <https://doi.org/10.1086/678293>
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). "Coh-metrix: Providing multilevel analyses of text characteristics". *Educational Researcher*, 40(5), 223–234. <https://doi.org/10.3102/0013189X11413260>
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). "Coh-Metrix: Analysis of text on cohesion and language". *Behavior Research Methods, Instruments, and Computers*, 36(2), 193–202. <https://doi.org/10.3758/BF03195564>
- Gulzar, M. A., Buriro, G. A., & Charan, A. A. (2017). "Investigating the effects of rubrics on assessment of writing tasks". *International Research Journal of Arts and Humanities*, 45(45), 191–206.
- Hughes, A. (1989). Testing for language teachers / Arthur Hughes. In *Idiom*. New York: Cambridge University Press.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, California: Sage Publications Inc.
- Lei, C., Man, K. L., & Ting, T. O. (2014). "Using Coh-Metrix to analyse writing Skills of students : A Case Study in a technological common core course". *Proceedings of the International MultiConference of Engineers and Computer Scientists 2014. Vol II*, 12–14.
- Matthews, J., & Wijeyewardene, I. (2018). "Exploring relationships between automated and human evaluations of L2 texts". *Language Learning and Technology*, 22(3), 143–158.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). "Automated evaluation of text and discourse". *The Elementary School Journal*, 115(2), 210–229
- Moskal, B. M. (2000). "Scoring rubrics: What, when and how?" *Practical Assessment, Research and Evaluation*, 7(3), 2000–2001.
- Ngadiso. (2013). "Assessing the students' writing skill". *Reforming Paradigms towards More Integrated Education*, 148–154.
- Quispesaravia, A., Perez, W., Cabezudo, M. S., & Alva-Manchego, F. (2016). "Coh-Metrix-Esp: A complexity analysis tool for documents written in Spanish". *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 4694–4698.
- Tortorelli, L. S. (2019). "Beyond first grade: examining word, sentence, and discourse text factors associated with oral reading rate in informational text in second grade". *Reading and Writing*, 33(1), 143–170. <https://doi.org/10.1007/s11145-019-09956-5>
- Zedelius, C. M., Mills, C., & Schooler, J. W. (2018). "Beyond subjective judgments : Predicting evaluations of creative writing from computational linguistic features." *Behavior Research Methods*, 51(2), 879–894. <https://doi.org/10.3758/s13428-018-1137-1>.