Does Automated Writing Evaluation (AWE) Improve Students' Writing? Focus on Technical Aspects and Readability

Luqmanul Hakim Zulkornain*, Azman Che Mat, and Nurul Ajleaa Abdul Rahman

***Abstract*—Automated Writing Evaluation (AWE) software has gained popularity in the field of writing related research. Most studies focused on perception and acceptance of AWEs and improvements in writing. However, this study was carried out based on the technicality of the writing which are often discussed as readability. The objectives of the study are to investigate the improvement in writing and to compare the results between the control and the experimental groups. The study employed a descriptive research design with two groups (control (n=49) and experimental n=72)) undergoing two tests (pre-test and post-test). The gap between these tests was 10 weeks where the control group went on the traditional teaching and learning method while the experimental group were exposed and trained to use AWE. Using the Flesch Reading Ease Scale and the Flesch-Kincaid Grade Level, the readability of the writings as well as the grade level required to understand the writings were tested. The findings revealed that the improvement of the readability in the experimental group fared slightly less compared to the control group. However, it is still safe to conclude that AWE does help to improve writings.**

***Index Terms*—Automated Writing Evaluation (AWE), readability, Flesch-Kincaid Grade Level, Flesch Reading Ease, writing**

I. Introduction

In recent years, Automated Writing Evaluation (AWE) software has gained popularity in the field of writing related research. AWE software provides computer-generated feedback for writing and users utilize them in reviewing and assessing their writings [1], illustrating their utilization as substitutes, assisting users to improve their writing. One of the most popular software is Grammarly. This AWE software is available on most digital platforms and has the potential to develop and improve writing abilities.

Grammarly has the potential to be utilized in writing classes. It is promoted as a simple application with the ability to assist students and academicians in their writing by providing suggestions in correcting spelling, grammar, and punctuation problems with detailed and useful comments. These could result in better writing, improving most aspects such as clarity, readability, and accuracy [2].

Writing is one of the productive skills of a language. It is crucial as it is one of the methods to convey meaning or ideas. Writing process requires instructors and learners to communicate for monitoring and providing relevant helpful feedback [3]. Writing can be assessed in terms of how ideas are conveyed and in terms of technicalities.

In writing, the technicalities; lexical density and length are usually discussed as readability. Measured in scales such as The Gunning Fog Index, The SMOG Index, as well as Flesch Kincaid Grade Level, and Flesch Reading Ease, the readability in writing represents how hard or how easy it is for a writing to be understood.

The Covid-19 Pandemic has affected every aspect of human lives. Countries had exercised the imposition of Movement Control Order (MCO) which had greatly limited movements and interactions to curb the spread of the disease. This imposition has affected the government and sectors, business, education, and lives in general [4–6]. The education was hit hard by the gruesome reality that classes could not be conducted as usual. One of the approaches taken was the implementation of Online Distance Learning (ODL), shifting physical learning to virtual learning environments [7]. Although ODL (also formerly known as Distance Education (DE)) has been discussed since the late 80's [8] and has been implemented since the 1700s [9], the rapid implementation in 2020 shocked those involved as most were not ready and did not have time to adjust but had to just embrace ODL [10]. ODL is not without flaw and one of the challenges in the implementation of ODL is barriers in communication which include understanding roles, anxiety, comprehension, and technicalities [11].

The pandemic has proven to be discouraging as both students and instructors lose the opportunities to engage in an authentic traditional face to face monitoring session and this also affects the teaching and learning of writing. Upon the observation of the features of AWE, the software could possibly be an alternative to substitute the role of instructors in the teaching and learning of writing.

II. Problem Statement

Writing is one of the productive skills in a language. It is crucial as it is one of the methods to convey meaning or ideas. The writing process requires instructors and learners to communicate for monitoring and providing relevant helpful feedback [3]. Writing can be assessed in terms of how ideas are conveyed and in terms of technicalities.

Due to Covid-19 Pandemic shifting physical learning space to virtual, it is impossible to engage all the learners in writing exercises, hindering guidance for the essential processes of writing which are to perceive, break down, and express a thought into a section or exposition [12]. This is where Automated Writing Evaluation (AWE) software may be useful as they do not only analyze writing mistakes but also provide quick feedback. Therefore, a good AWE software is required for error identifications and recommendations on changes to be made as well as why they are required. Grammarly, one of the

Manuscript received March 8, 2023; revised April 27, 2023; accepted June 27, 2023.

Luqmanul Hakim Zulkornain, Azman Che Mat, and Nurul Ajleaa Abdul Rahman are with Akademi Pengajian Bahasa, Universiti Teknologi MARA Terengganu, Dungun, Terengganu, Malaysia.

*Correspondence: luqman562@uitm.edu.my (L.H.Z.)

many AWEs provide assistance in writing in which it detects errors and provides feedback to improve sentences and structures with explanation. It can be integrated into writing software and applications making it easy to use [13].

Most studies such as [1, 14–16] only look into perception and acceptance of AWEs. Therefore, in ensuring the effectiveness of the utilization of AWE software as an assistant to improve writing, there is a need for studies to be conducted to support and prove its effectiveness. While most studies on improvements in writing such as [17–19] were only focused on error detection and accuracy, the study however, focused on the technicality of the writing which include lexical density and length which are often discussed as readability, measured in scales representing how hard or how easy it is to understand the writing. The study utilizes Flesch Kincaid Grade Level, and Flesch Reading Ease in examining the readability of students' writings.

The decision to investigate technicalities instead of other factors stemmed from the idea of the application of the Flesch-Kincaid Grade Level in the U.S. Navy in 1976 [20, 21]. The readability of manuals in the military were tested to make sure that they were comprehensible by those serving at that time as soldiers drafted came from various educational background.

To address the problem discussed in the previous section, the research aims to achieve two objectives which are outlined as the followings:

1) To investigate the improvement in students writing after being exposed to AWE.
2) To compare improvements between control and experimental groups.

In achieving these objectives, two research questions were formulated in which the first is to look at the kind of improvement in writing which can be measured by technicalities in writing (readability), and the second is the comparison between the control group and the experimental group. Hence, the research questions are as follows:

1) What kind of improvements in writing can be observed after students were exposed to AWE?
2) Does exposing students to AWE yield better improvements in writing compared to students exposed to the traditional method of teaching?

III. Literature Review

A. Writing

Writing is a challenging talent [22], and this is supported by Faller [23] in which it was mentioned that "writing is such a challenging endeavor that requires a great cognitive and linguistic abilities". One of the most challenging skills to be mastered by English as a Second Language (ESL) learners is the writing skill and this is due to the inadequacy and unsuitable language acquisition techniques as well as being unaware of the acceptable ESL writing methods [24]. Various writing techniques require the learners and the educators to collaboratively choose the best technique suiting the learners. Learners have to be equipped with suitable techniques to improve their writing and be taught on how to successfully employ it to boost their writing skills.

The observation of the process of writing and coming up with feedback do require time and effort and it is somehow abstract [3]. ESL learners usually make mistakes in writing, and these could be in spelling, which could affect meaning of words, and syntax, which could affect the meaning of the whole sentence [25–27]. These mistakes are somehow usually overlooked as they accidentally occur [28].

Writing can be assessed in terms of how ideas are conveyed and in terms of technicalities. Technicalities in writing such as lexical density and length are often discussed as readability. The readability of a writing is usually measured in scales representing how hard or how easy it is to understand the writing. There are several known scales in the field of readability such as The Gunning Fog Index, The SMOG Index, as well as Flesch Kincaid Grade Level, and Flesch Reading Ease.

B. Readability/Technical

1) Flesch Reading Ease

Designed by Rudolph Flesch in 1948 [29, 30], the Flesch Reading Ease was developed to measure the context of a writing. The difficulty to understand a writing is indicated by the score resulting from calculation done on the piece of writing based on a formula which is based on the length of word, length of sentence, word form, as well as syllables [31].

As presented by researches, the score for Flesch Reading Ease ranges from 0 to 100 [20, 29, 30]. Writings with scores ranging from 100 to 70 are considered "easy" to understand while scores 60 to 70 are considered as "standard". Scores from 60 to 0 are considered "difficult" and the most "difficult" would be score 0.

2) Flesch-Kincaid Grade Level

The Flesch-Kincaid Grade Level was developed in 1976 and was commissioned by the U.S. Navy [20, 21]. It was developed to measure the readability of the military manual. The Flesch-Kincaid Grade Level recalibrates the Flesch Reading Ease score into U.S. grade school level [29].

DuBay [20] and Derar *et al*. [31] illustrated that from Flesch-Kincaid Grade Level, the most standard writings are those of the 8th grade based on the U.S. grade level. These writings are easily comprehensible by those aged 13 to 15.

3) Automated Writing Evaluation (AWE)

Automated Writing Evaluation (AWE) software provides computer-generated feedback for writing. They are useful due to their ability to analyze writing mistakes and quick feedback. AWE software users utilize them for reviewing and to assess their writings [1]. One of the most popular among these is Grammarly. This AWE software is available on most digital platforms and has the potential to develop and improve writing abilities.

Grammarly detects errors and provides feedback to improve sentences and structures with explanation. It provides suggestions to improve spelling, grammar, and punctuation. These suggestions come with detailed and useful comments thus help in improving clarity, readability, and accuracy of a writing [2]. Grammarly can be integrated into writing software and applications making it easy to use [13].

Most studies related to the use of AWEs in writing, only look into perception and acceptance of AWEs [1, 14–16]. Hence, the impact of the AWE is not really discussed. Therefore, in ensuring that the exposure to and the utilization of AWE software as an assistant to improve writing actually improves students' writings, there is a need for studies to be conducted to support and prove its' effectiveness.

Most studies on improvements in writing were only focused on error detection and accuracy [17–19]. This is appropriate when it comes to only spelling and grammar accuracy. In terms of readability or reading ease, which is closely related to the complexity of the writing itself, readability test also has to be discussed. The current study looks into the technicality of the writing which includes lexical density and length which are often discussed as readability, measured in scales representing how hard or how easy it is to understand the writing. The study utilizes Flesch Kincaid Grade Level to determine the maturity of the writing hence its complexity, and Flesch Reading Ease in examining the readability of students' writings.

IV. Methodology

A. Research Design

The design of the study is descriptive in nature where it is aimed to identify characteristics, frequencies, and trends for the results of the two categories (control group and experimental group) of students involved in the study.

The study observed the improvements in writing of the participants for the two groups in which the experimental group was exposed and trained to use AWE (Grammarly) while the control group was only exposed to the traditional method of teaching writing.

Grammarly was chosen due to its popularity and availability on most digital platforms where it is promoted as a simple application or software which can help users improve their writings by providing suggestions in correcting spelling, grammar, and punctuation problems with detailed and useful comments.

B. Participants

The participants of the study consists of 121 students from four universities in Malaysia namely Universiti Teknologi Mara Cawangan Terengganu (UiTMCTKD), Universiti Malaya (UM), Universiti Malaysia Kelantan (UMK), and Universiti Malaysia Pahang (UMP). The number of participants is broken down and illustrated in the Table I.

TABLE I: Participants of the Study

University	UiTMCTKD	UM	UMK	UMP	Total
Experimental Group	32	16	15	9	72
Controlled Group	19	11	12	7	49
Total					121

To avoid bias, the participants chosen were not those enrolling in English language related streams such as those majoring in English language studies, English literature, English for communication, and Teaching English as Second Language (TESL). This is to test the effectiveness of AWE, in this case, Grammarly, on students in general. The participants at the time of when the study was conducted were all enrolling in writing class.

C. Data Collection

The data was based on two data sets which are pretest writings and post-test writings. In between tests, students in the experimental group were exposed and trained to use Grammarly, for 10 weeks. The training required students to utilize the software for all their writings and learn as much from the feedback provided by the software. The control group however, underwent traditional teaching and learning method without the intervention of any AWEs.

D. Instrument

The instruments employed in the study were pretests and post-tests questions. In these tests, participants were required to write a short essay for each test. The topic selected for the essay was on "Admired Person" which is relatable to most people. The reason for the selection is to avoid bias. This is similar to what were done by [32, 33].

E. Analysis

1) Tool

The tool employed for the analysis was a readability test tool. Readability tools are in abundance and can even be found in standard word processing software such as the Microsoft Word [20, 29]. This study however, employed the Readability Test and Improve Tool available on Online-Utility.Org website.

The selection for this tool is due to its ability to provide readability analysis for many readability tests and scores (SMOG, The Automated Readability Index (ARI), Gunning-Fog, Coleman-Liau, Flesch Reading Ease, and Flesch-Kincaid Grade Level), all at once. This is illustrated in Fig. 1.

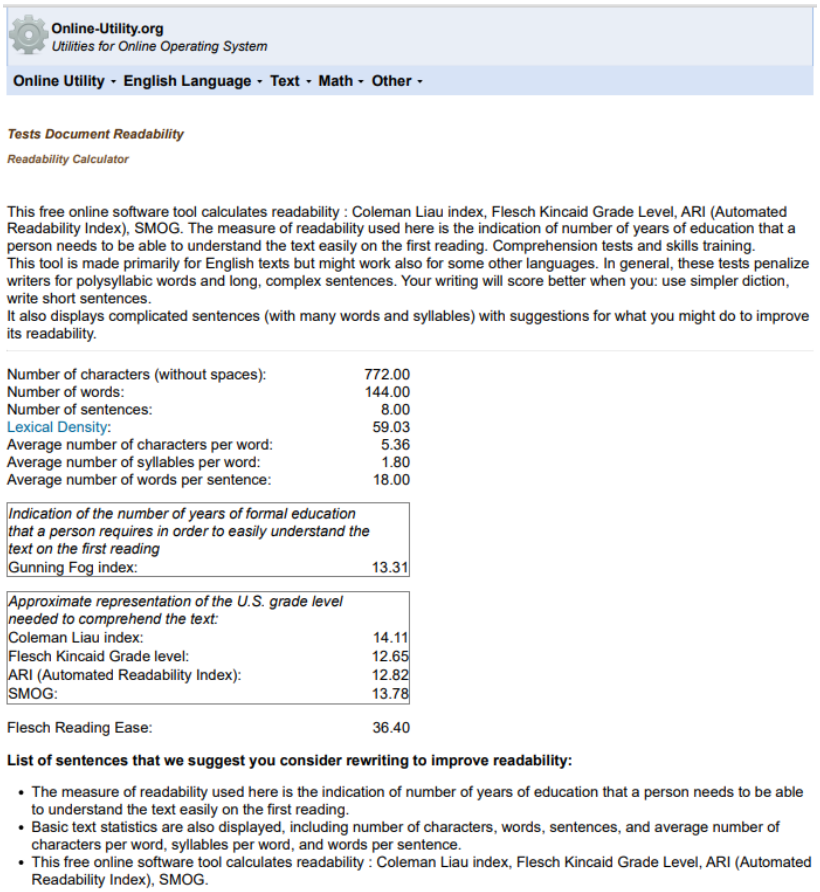
The study however, only looks at Flesch Reading Ease, and Flesch-Kincaid Grade Level. Flesch Reading Ease provides reading ease scores determining how easy it is to understand a piece of writing. This is measurable via the formula:

$$206.835 - 1.015 \times \left(\frac{total\ words}{total\ sentences}\right) - 84.6 \times \left(\frac{total\ syllables}{total\ words}\right)$$

The score of the writing based on the calculation determines how easy it is to understand the writing. Higher scores indicate that the writing is easier to comprehend. This is illustrated in the following Table II.

TABLE II: Flesch Reading Ease Score [20]

Reading Ease Score	Style Description	Estimated Reading Grade	Estimated Percent of U.S. Adults (1949)
0 to 30:	Very Difficult	College graduate	4.5
30 to 40:	Difficult	13th to 16th grade	33
50 to 60:	Fairty Difficult	10th to 12th grade	54
60 to 70:	Standard	8th and 9th grade	83
70 to 80:	Fairly Easy	7th grade	88
80 to 90:	Easy	6th grade	91
90 to 100:	Very Easy	5th grade	93


Online-Utility.org
Utilities for Online Operating System

Online Utility - English Language - Text - Math - Other -

Tests Document Readability

Readability Calculator

This free online software tool calculates readability : Coleman Liau index, Flesch Kincaid Grade Level, ARI (Automated Readability Index), SMOG. The measure of readability used here is the indication of number of years of education that a person needs to be able to understand the text easily on the first reading. Comprehension tests and skills training.
This tool is made primarily for English texts but might work also for some other languages. In general, these tests penalize writers for polysyllabic words and long, complex sentences. Your writing will score better when you: use simpler diction, write short sentences.
It also displays complicated sentences (with many words and syllables) with suggestions for what you might do to improve its readability.

Number of characters (without spaces):	772.00
Number of words:	144.00
Number of sentences:	8.00
Lexical Density:	59.03
Average number of characters per word:	5.36
Average number of syllables per word:	1.80
Average number of words per sentence:	18.00

Indication of the number of years of formal education that a person requires in order to easily understand the text on the first reading	
Gunning Fog index:	13.31

Approximate representation of the U.S. grade level needed to comprehend the text:	
Coleman Liau index:	14.11
Flesch Kincaid Grade level:	12.65
ARI (Automated Readability Index):	12.82
SMOG:	13.78

Flesch Reading Ease: 36.40

List of sentences that we suggest you consider rewriting to improve readability:

- The measure of readability used here is the indication of number of years of education that a person needs to be able to understand the text easily on the first reading.
- Basic text statistics are also displayed, including number of characters, words, sentences, and average number of characters per word, syllables per word, and words per sentence.
- This free online software tool calculates readability : Coleman Liau index, Flesch Kincaid Grade Level, ARI (Automated Readability Index), SMOG.

Fig. 1. Readability test tool [34].

Flesch-Kincaid Grade Level determines the grade level or the maturity of the writer based on the writing. These grade levels are based on The U.S. school grade level. Even though the grade levels are based on The U.S. school grade level, it can be used across the globe and is one of the well-received tools in determining the maturity of a writing that it can be found as standard in mainstream writing software such as the Microsoft Word. The formula to determine the grade level is as the following:

$$0.39 \times \left(\frac{total\ words}{total\ sentences}\right) + 11.8 \times \left(\frac{total\ syllables}{total\ words}\right) - 15.59$$

Higher grade level depicts higher maturity and proficiency in writing while lower grade level depicts lower maturity and proficiency in writing. These depictions can be observed in the following Fig. 2, which comes with examples of writing for several grade level.

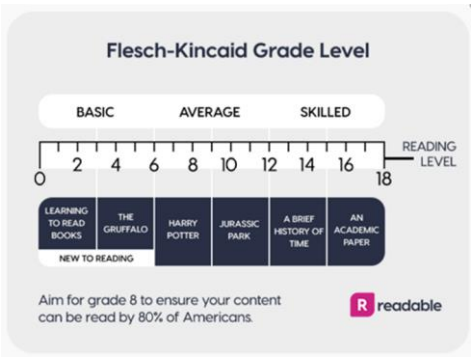


Fig. 2. Flesch-Kincaid Grade Level [35].

The readability results (Flesch Reading Ease Score and Flesch-Kincaid Grade Level) for the pretests and post-tests were compared. Comparison was also made between the control and the experimental groups.

V. Findings and Discussion

The findings illustrate an interesting revelation in the readability tests conducted. The study only focused on two readability tests which are the Flesch-Kincaid Grade level and the Flesch Reading Ease. Although both tests may appear the same in their names, they indicate different measurements as Flesch Reading Ease measures how easy it is to comprehend a writing based on the scores given to the writing while Flesch-Kincaid Grade level measures the maturity or the grade level of the writer based on The U.S. school grade level.

A. Flesch Reading Ease

The average Flesch Reading Ease scores of both pretest and post-test for the experimental group are depicted in Table III.

TABLE III: Average Flesch Reading Ease Score

Test	Average Flesch Reading Ease Score			
	Experimental Group		Control Group	
	Score	Category	Score	Category
Pre	65.66	Standard	66.10	Standard
Post	67.37	Standard	67.40	Standard

Both pretest and post-test average scores for the experimental group still fall under the "standard" category where the writings are neither easy nor hard to understand. The pretest average score is 65.66 while there is an improvement of 1.71 when compared to the average post-test score (67.37). The improvements here illustrate that the average writing has an improved readability although both results still fall under the same category.

Closer investigation on the scores reveals that 50% (n = 36) of the participants in the experimental group have shown improvements in their Flesch Reading Ease score. 36.11% (n = 26) shows degradation, while 13.89% (n = 10) remains stagnant. This is observable in Table IV.

Similar to the experimental group, the average scores for the control group still fall under the standard category indicating that the writings for this group are neither too hard nor too easy to understand. The data illustrates that for the pretest, the average score is 66.10 and an improvement of 1.30 can be seen when it is compared to the post-test score (67.40). These slightly differ from the experimental group where its pretest average score is 65.66 and the average post-test score is 67.37. The difference in the average score is better for the experimental group which is at 1.71 compared to 1.30 for the control group. This indicates that although the average of improvement in writing is better in the control group (avg. control group post-test score: 67.40 vs. avg. experimental group post-test score: 67.37), the quality of the improvement is better in the experimental group (avg. control group score diff: 1.71 vs. avg. experimental group score diff: 1.30).

Upon closer investigation, when the pretest results and the post-test results were compared, it can be observed in Table IV that 57.14% (n = 28) of the control group manifested improvement in their score while 42.86% (n = 21) manifested degradation in grade. None of the participants was showing stagnant results. Compared to the results for the experimental group, the results for the control group are proven to be better as it can be found that the percentage of those having their score improved is better (control group: 57.14% vs. experimental group: 50.00%) although the percentage of those experiencing degradation in score is higher (control group: 42.86% vs. experimental group: 36.11%).

TABLE IV: PERCENTAGE OF FLESCH READING EASE SCORE CHANGES

	Experimental Group		Control Group	
	Participants		Participants	
Score	Number	(%)	Number	(%)
Improve	36	50%	28	57.14%
Degrade	26	36.11%	21	42.86%
Stagnant	10	13.89%	0	0%
Total	72		49	

As illustrated in Table V, the average improvement in the Flesch Reading Ease Score for experimental group is 7.28 while the average score degradation is −5.36. It can be observed from the data that the average in the improvement is greater compared to the degradation in the Flesch Reading Ease score by 1.92. This interprets that in average, AWE does improve the readability of writing among students.

For the control group, although more participants manifest improvement, Table V reveals that the average score improvements are slightly lower than degradation. The average improvement in the Flesch Reading Ease Score is 9.83 while the average score degradation is −10.07. The difference here is 0.24 where degradation score exceeds improved score. Comparing this to the experimental group, it can be said that the experimental group performs better as the difference in the average improved grade compared to degradation is at 1.92 with the former being greater than the later

TABLE V: AVERAGE SCORE CHANGE DIFFERENCE

Experimental Group		Control Group	
Score Change	Average	Score Change	Average
Improve	7.28	Improve	9.83
Degrade	−5.36	Degrade	−10.07

B. Flesch-Kincaid Grade Level

Table VI illustrates the average grade levels of the pretest and post-test for the experimental group. For the pretest, the average for the grade level is 8.40. There has been a drop in the average grade level for the experimental group as it can be observed from the data that the average grade level for the post-test is 7.92 which is lower than the pretest average results. The difference between the average Flesch-Kincaid Grade Level for the experimental group is 0.48.

Compared to the experimental group, the average grades for the pretest and the post-test only display minute differences with the pretest for the control group at 8.41 (8.40 for experimental) and 7.96 (7.92 for experimental) for the post-test.

TABLE VI: AVERAGE FLESCH-KINCAID GRADE LEVEL

	Experimental Group		Control Group	
Test	Average	Skills	Average	Skills
Pre	8.40	Average	8.41	Average
Post	7.92	Average	7.96	Average

Further investigation yielded more comprehensive findings resulting in the drop in grade discussed previously for the experimental group. Of the 72 participants in the experimental group, not all of them experienced grade drop. The number of those experiencing grade drop is 33 representing 45.83% of the population. 38.89% (n = 28) experienced improvements in grade while 15.28% (n = 11) remained stagnant. This can be observed in Table VII.

Similar to the experimental group, Table VII also illustrates that more participants experienced degradation in grade level. However, none remained stagnant. It can be observed that more than half (57.14%, n = 28) of the participants experienced degradation in grade level while 42.86% (n = 21) have their grade level improved. However,

the percentage suggests that both improvement and degradation are more prevalent in the control group where the percentages of improvement and degradation of the experimental group are respectively 38.89% and 45.83%, which are lower than those of the control group.

TABLE VII: PERCENTAGE OF FLESCH-KINCAID GRADE LEVEL CHANGES

	Experimental Group		Control Group	
	Participants		Participants	
Grade	Number	(%)	Number	(%)
Improve	28	38.89%	21	42.86%
Degrade	33	45.83%	28	57.14%
Stagnant	11	15.28%	0	0%
Total	72		49	

The improvements and degradations in grade reveal another interesting result. For the experimental group, the average grade degradation is almost double of the average grade improvement. It can be observed from Table VIII that the average improvement is 0.94 while the average degradation is −1.83.

For the control group, the average improvement is 2.11 while the average degradation is −2.38.

Although the degradation of grade is found to be more prominent, it does not necessarily mean that the writings have not been improved at all as Flesch-Kincaid Grade Level only indicates the grade level of the writers.

TABLE VIII: AVERAGE GRADE LEVEL CHANGE DIFFERENCE

Experimental Group		Control Group	
Grade Level Change	Average	Grade Level Change	Average
Improve	0.94	Improve	2.11
Degrade	−1.83	Degrade	−2.38

In the comparison of the experimental and the control group, both average improvement and degradation are greater in the control group. The average for improvement is more than double (2.11(control) to 0.94(experimental)) while although not as much as the comparison in improvement, the degradation still shows an increase (2.38(control) to 1.83(experimental)).

C. Revisiting the Research Questions

1) What kind of improvements in writing can be observed after students were exposed to AWE?

From the results, it can be found that there are two changes (improvements and degradations) which can be observed when it comes to readability of students writing after being exposed to AWE. The first is the readability score and the later is the grade level.

The results revealed that in general, the readability score of the participants improved in the post-test. However, the grade level of the participants degraded. This illustrates that the writings were easier to comprehend despite their decreasing grade level in writing.

2) Does exposing students to AWE yield better improvements in writing compared to students exposed to the traditional method of teaching?

Comparing the experimental group and the control group yielded an interesting finding. The average reading ease score for both pretest and post-test are better in the control group. The improvement in the percentage of those who have improved reading ease score can also be observed to be better in the control group. When it comes to the quality of improvement, the experimental group performed better.

In terms of grade level, the control group manifested better results as more participants have had their grade level improved with the average for the post-test grade level being slightly higher than of the experimental group. The quality of the improvement in grade level, however, was better in the experimental group with the average grade jump being almost one grade up.

These differences, however, are not too prevalent and can still be susceptible to other factors such as imbalanced participants making up both groups and the topic chosen which was too generic.

VI. CONCLUSION

From the findings, it can be observed that both groups (experimental and control) experienced improvement in their Flesch Reading Ease scores depicting that the writings of these students were easier to be understood after the experiment. The percentage of those experiencing improvement in reading ease score has also improved in both groups. The average value of the score however tells a slightly different story where the experimental group has shown that its average score improvement is better than their average score degradation while for the control group, although the average score improvement is superior to the experimental group's average score degradation, it is still slightly being overwhelmed by its own average score degradation. The average reading ease score for both pretest and post-test are better in the control group. The improvement in the percentage of those who have improved reading ease score can also be observed to be better in the control group.

For Flesch-Kincaid Grade Level, it can be concluded that in average, both control and experimental groups experienced degradation in grade level with the experimental group experiencing more degradation. Despite the degradation in the average, there were still improvements in the grade level among the participants. The improvements in grade level in the control group doubled that of the experimental group. However, when it comes to degradation, the control group also exceeded the experimental group.

The application of AWE does help with improving the writing of students when it comes to make it easier to understand. However, the improvement is not as good when compared to the traditional method where students were just being monitored and receiving feedback from instructors. These are reflected in the degradation of grade level as well as the score for reading ease. Despite all the differences, it is still safe to conclude that AWE does help to improve writings as the differences found were minute and not too prevalent.

The research however only takes into account the aspect of

readability where it was only constrained to lexical density and length. Other aspects such as the quality of sentence structure, semantic, and pragmatics were discarded. Further studies pertaining to AWEs and writing can be conducted taking into account errors in sentence structure as well as semantics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Azman Che Mat and Nurul Ajleaa contributed in the literature section regarding AWE while Luqmanul contributed in the Readability section. The authors teamed up in the conceptualization and data collection phases. Luqmanul was responsible for the analysis, pre-writing and the completion of the writing while Azman and Nurul Ajleaa were responsible for the additional information required for the study. All authors had approved the final version.

FUNDİNG

This study was conducted with funding from Professorial Research Grant under file number: 600-RMC/GPP5/3(002/2021).

REFERENCES

[1] R. ONeill and A. Russell, "Stop! Grammar time: University students' perceptions of the automated feedback program grammarly," *Australasian Journal of Educational Technology*, vol. 35, no.1, pp. 42–56, 2019.
[2] Beyond grammar and spelling. *Grammarly*. [Online]. Available: https://www.grammarly.com/.
[3] V. L. Fei and J. Phua, "Teaching writing with language feedback technology," *Computers and Composition*, vol. 54, pp. 1–18, 2019.
[4] Mathew, V. Nity, and E. Chung, "University students' perspectives on Open and Distance Learning (ODL) implementation amidst COVID-19," *Asian Journal of University Education*, vol.16, no.4, pp. 152–160, 2020.
[5] Shahzad, S. Khuram, *et al*., "Impact of virtual teaching on ESL learners' attitudes under COVID-19 circumstances at post graduate level in Pakistan," *English Language Teaching*, vol. 13, no. 9, pp. 1–9, 2020 .
[6] P. D. MacIntyre, T. Gregersen, and S. Mercer, "Language teachers' coping strategies during the Covid-19 conversion to online teaching: Correlations with stress, wellbeing and negative emotions," *System*, vol. 94, pp. 1–13, 2020.
[7] S. P.-L. Sim, H. P.-K. Sim, and C.-S. Quah, "Online learning: A post COVID-19 alternative pedagogy for university students," *Asian Journal of University Education*, vol. 16 no. 4, pp. 137–151, 2021.
[8] S. N. S. Allam *et al*., "Online distance learning readiness during Covid-19 outbreak among undergraduate students," *International Journal of Academic Research in Business and Social Sciences*, vol. 10, no. 5, pp. 642–657, 2021.
[9] B. Aras, "From distance education to open and distance learning: A holistic evaluation of history, definitions, and theories," *Handbook of Research on Learning in the Age of Transhumanism. IGI Global*, pp. 252–273, 2019.
[10] Brief, Policy, "Education during COVID-19 and beyond," United Nations, pp. 1–26, 2020.
[11] I. Aytekin and F. Dabaj, "Communication barriers in distance education," *The Turkish Online Journal of Educational Technology*, vol. 2, no. 4, 2003.
[12] Y. Moh, "Learning from students' experiences in writing paragraph," *Metathesis: Journal of English Language, Literature, and Teaching*, vol. 3, no. 2, pp. 188–202, 2019.
[13] K. Svetlana, "Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study," *Assessing Writing*, vol. 44, 2020.
[14] S. Aysel and Z. Bilki, "Voluntary use of automated writing evaluation by content course students," *ReCALL*, vol. 33, no. 3, pp. 265–277, 2021.
[15] R. O'Neill and A. Russell, "Grammarly: Help or hindrance? Academic learning advisors' perceptions of an online grammar checker," *Journal of Academic Language and Learning*, vol. 13, no. 1, pp. A88–A107, 2019.
[16] E. K. Ambarwati, "Indonesian university students' appropriating Grammarly for formative feedback," *ELT in Focus*, vol. 4, no. 1, pp. 1–11, 2021.
[17] P. G. Lorena and X. S. Calero, "Automated writing evaluation tools in the improvement of the writing skill," *International Journal of Instruction*, vol. 12, no. 2, pp. 209–226, 2019.
[18] L. Stephanie, M. Mehrzad, and M. Rahimi, "Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement," *Computer Assisted Language Learning*, vol. 35, no. 4, pp. 605–634, 2022.
[19] B. Hamed, R. Vaseghi, and R. Gholami, "The effect of diary writing on EFL college students' writing improvement and attitudes," *International Conference on Languages, Literature and Linguistics IPEDR*. vol. 26, pp. 143–147, 2011.
[20] W. H. DuBay, "The principles of readability," *Costa Mesa*, CA: Impact Information, 2004.
[21] J. Pranay and J. C. MacDermid, "Assessing reading levels of health information: uses and limitations of flesch formula," *Education for Health*, vol. 30, no. 1, pp. 84–88, 2017.
[22] M. M. Mulianing, "Graphic organizers to improve students' writing on recount paragraphs," *Metathesis: Journal of English Language, Literature, and Teaching*, vol. 2, no. 2, pp. 211–221, 2018.
[23] J. M. V. Faller, "Grammarly investigation into EFL writing issues involving Omani learners," *International Journal of Language & Linguistics*, vol. 5, no. 3, pp. 165–174, 2018.
[24] C. Yang, "How Chinese beginning writers learn English writing: A survey of writing strategies," *Journal of Educational and Social Research*, vol. 3, no. 1, pp. 9–18, 2013.
[25] T. N. Fitria, "Error analysis found in students' writing composition in simple past tense of recount text," *ENGLISH FRANCA: Academic Journal of English Language and Education*, vol. 4, no. 2, pp. 141–160, 2020.
[26] T. N. Fitria, "Spelling Error Analysis in Students'writing English composition," *Getsempena English Education Journal*, vol. 7, no. 2, pp. 240–254, 2020.
[27] P. Iwan and M. Farida, "Online grammar checkers and their use for EFL writing," *Journal of English Teaching, Applied Linguistics and Literatures (JETALL)*, vol. 2, no. 2, pp. 67–76, 2019.
[28] T. N. Fitria, "Error analysis found in students' writing composition of simple future tense," *ELS Journal on Interdisciplinary Studies on Humanities*, vol. 1, no.3, pp. 240–251, 2018.
[29] Z. Mostafa and P. Heydari, "Readability of texts: State of the art," *Theory & Practice in Language Studies*, vol. 2, no. 1, pp. 43–53, 2012.
[30] G. Vijay, "Readable writing: Measures and significance for teaching & teacher," *International Journal of Multidisciplinary Educational Research*, vol. 4, no. 3, pp. 1–14, 2015.
[31] . Derar, A. Othman, and A. Eleyan, "Enhancing software comments readability using flesch reading ease score," *Information*, vol. 11, no. 9, pp. 1–25, 2020.
[32] P. Anjali, P. Rout, and B. K. Swain, "An action research study on improving seventh standard students' creative writing skill in English," *IOSR Journal of Humanities and Social Science*, vol. 22, no. 6, pp. 13–21, 2016.
[33] T. N. Kham and M. Nikolov, "How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing," *The Asia-Pacific Education Researcher*, vol. 31, no. 6, pp. 767–779, 2022.
[34] *Tests Document Readability*. [Online]. Available: https://www.online-utility.org/english/readability_test_and_improve.jsp.
[35] (2019). *Readability Formulas*. [Online]. Available: https://readable.com/features/readability-formulas/