Akademika 94(2), 2024: 474-491
https://doi.org/10.17576/akad-2024-9402-27

Evaluating Automated Grammar Corrective Feedback Tools: A Comparative Study of Grammarly and QuillBot in ESL Expository Essays

Penilaian Alat Pembetulan Tatabahasa Secara Automatik: Kajian Perbandingan di antara Grammarly dan QuillBot untuk Esei Ekspositori Bahasa Inggeris sebagai Bahasa Kedua

NAJLAA' NASUHA MOHD RADIN, AIDA MUSTAPHA
& AILEEN FARIDA MOHD ADAM

ABSTRACT

The advent of artificial intelligence and the proliferation of automated grammar feedback applications have garnered great interest among ESL learners as tools to facilitate language acquisition. While ample studies have examined the utility of applications like Grammarly and Quillbot, scarce research compares their effectiveness in identifying and classifying errors in Malaysian ESL student writing samples. This study aimed to conduct such a comparative analysis using expository essays authored by Malaysian ESL students. This study employs a descriptive quantitative approach to collect data and conduct data analysis. Five writing samples were examined using both applications to ascertain the frequencies of errors flagged and categorised mistakes based on James' (1998) error classification schemata. Results demonstrated that overall, Grammarly detected more errors compared to Quillbot. Additionally, both applications recognised substantially more grammatical and substance inaccuracies relative to other error types like lexical, syntactic, or semantic issues. Grammarly provided detailed descriptions and suggestions of each error identified, while Quillbot only highlighted the errors with brief explanations. These findings suggest both tools can meaningfully supplement ESL learners in their language learning process. However, further investigations into their respective strengths and limitations are merited given the nuances observed. Overall, this exploratory study highlights the promise of automated writing evaluation to enable self-directed editing to enhance the language learning process among ESL learners.

Keywords: Automated writing evaluation; Error analysis; ESL learners; Language learning tools; Error classification

ABSTRAK

Kemunculan kecerdasan buatan dan perkembangan penggunaan alat pembetulan tatabahasa automatik telah menarik minat para pelajar Bahasa Inggeris sebagai Bahasa Kedua (BI sebagai BK) untuk digunapakai sebagai alat pembelajaran bahasa Inggeris mereka. Walaupun banyak kajian dihasilkan berkaitan dengan penggunaan Grammarly dan QuillBot, kajian yang membandingkan keberkesanan mereka dalam mengenal pasti dan mengkategorikan kesilapan dalam sampel penulisan pelajar BI sebagai BK di Malaysia masih terhad. Kajian ini bertujuan untuk menilai perbandingan Grammarly dan Quillbot menggunakan esei ekspositori yang ditulis oleh pelajar BI sebagai BK di Malaysia. Kajian ini menggunakan pendekatan kuantitatif dalam mengumpul dan menganalisa sampel. Lima sampel diperiksa menggunakan kedua-dua platform untuk menentukan kekerapan kesilapan dikenal pasti dan mengkategorikan kesilapan berdasarkan klasifikasi kesilapan yang dicadangkan oleh James (1998). Hasil menunjukkan Grammarly mengenal pasti lebih banyak kesilapan berbanding QuillBot. Selain itu, kedua-dua alat mengenal pasti banyak kesalahan tatabahasa dan substansi berbanding jenis kesilapan lain seperti leksikal, sintaksis, atau semantik. Grammarly menerangkan dan mencadangkan secara terperinci setiap kesilapan yang dikenal pasti, sementara QuillBot hanya mengenal pasti kesilapan dengan penerangan yang ringkas. Penemuan menunjukkan kedua-dua alat ini dapat memberi sumbangan yang bermakna kepada pelajar BI sebagai BK dalam proses pembelajaran bahasa mereka. Namun, penyelidikan lanjut tentang kelebihan dan batasan aplikasi tersebut perlu dilakukan secara lebih mendalam. Secara keseluruhan, kajian ini menjamin bahawa alat pembetulan tatabahasa automatik memberi peluang kepada pelajar BI sebagai BK untuk melakukan pembetulan secara kendiri bagi meningkatkan pembelajaran mereka.

Kata Kunci: Penilaian penulisan automatik; Analisis kesilapan; Pelajar BI sebagai BK; Alat pembelajaran bahasa; Pengelasan kesilapan

INTRODUCTION

The main objective of ESL/EFL classrooms is to teach and guide language learners to be proficient in reading, writing, speaking, and listening in English. To guarantee that language learners can effectively communicate and convey their ideas, feelings, and thoughts in the target language, the two productive skills—writing and speaking—are crucial. However, there is no denying that there have always been significant obstacles and complexity in the way of learning written communication as a useful language ability (Nor Nadia Raslee et al., 2022). In addition to learning the language, students should also focus on improving their composition, spelling, grammar, flow, and organization of ideas into legible writing. (Hassan et al. 2021). To ensure that their written works are understandable and accurate, students are taught the principles of coherence, proper grammar, lexical structures, acceptable spelling, and appropriate language register choices in ESL writing lessons. This is done because the instructors recognise the need to integrate these skills while writing.

The development of digital technology has made it possible for both language teachers and students to benefit immensely from the integration of technology into language learning. Language learners have benefited from the introduction of several artificial intelligence-based writing tools, including Grammarly, QuillBot, Ginger, Hemingway Editor, and ProWriting Aid. These AI-powered resources can help students with a variety of tasks, including grammar correction, translation, text summarisation, sentence paraphrasing, and much more (Shadiev & Feng, 2023). Undoubtedly, these resources have aided many language learners in their quest to be better language learners.

Although numerous scholarly works emphasise the effectiveness of Grammarly in providing automated grammar feedback to learners (Soegiyarto et al. (2022); Moon (2021)), there is less studies pertaining to the utilisation of QuillBot. Prior research on QuillBot concentrated mostly on its functionality as a tool for paraphrasing and rewriting English prose (Reguig, & Mouffok, 2023; Raheem, et al., 2023). Apart from that, most of the studies carried out on the effectiveness of Grammarly and QuillBot only focused on the free versions of both applications.

Abundant studies have been conducted to assess the usefulness of Grammarly and QuillBot (John & Woll, 2020; Ambarwati, 2021 & Chui, 2022), but there is an absence of comprehensive research on how effectively Grammarly and QuillBot identify and classify errors in ESL expository essays, particularly among Malaysian ESL students. Hence, this study was conducted to better understand the strengths and limitations of these tools in an educational setting. The scope of this study focuses on the corrective feedback of expository essays with the aim of providing insights into the effectiveness of these tools in improving writing accuracy. Thus, below are two objectives of this study:

(1) To identify the total number of errors detected by Grammarly and QuillBot in ESL expository essays,
(2) To categorise the types of errors identified by both Grammarly and QuillBot in ESL expository essays, and

Based on the study objectives above, below are two research questions of this study:

(1) How many errors are identified by both Grammarly and QuillBot, and
(2) What type of errors do Grammarly and QuillBot identify most in the essays?

Akademika 94(2), 2024: 474-491
https://doi.org/10.17576/akad-2024-9402-27

LITERATURE REVIEW

CORRECTIVE FEEDBACK IN ENGLISH AS A SECOND LANGUAGE (ESL) ESSAY WRITING

English as a Second Language (ESL) essay writing is a complex task that poses challenges for second-language learners. Among the issues highlighted in past studies on ESL essay writing are accuracy, vocabulary, and coherence (Al Faruqy, 2022). The introduction of IR4.0 has transformed many aspects of life, including the education sector with automation and robotics systems (Fathiyah et al., 2022). In line with IR4.0, the use of technology in education is growing rapidly. The integration of digital technologies in education is revolutionising the way lessons are crafted in the 21st century (Chen, 2022). The use of digital technologies in essay writing has shown to positively impact both the written texts and the writing process (Mahapatra, 2024). These digital tools, platforms, and collaborative opportunities support effective organisation and presentation of ideas in ESL writing.

Automated corrective feedback tools can significantly enhance students' writing skills and accuracy. These tools provide immediate and consistent feedback, allowing students to make timely revisions and practice self-editing skills (Shadiev & Feng, 2023; Ranalli, Link, & Chukharev-Hudilainen, 2017). They can handle large volumes of work efficiently, which is especially useful in contexts with large class sizes and limited instructor availability (Wilson & Roscoe, 2020). Moreover, automated feedback encourages student independence by helping learners take initiative in identifying and correcting their errors without relying solely on the instructor.

In the ESL writing classroom, numerous studies have investigated the effects of the various types of written corrective feedback on student writing. These studies include teacher feedback (Wondim, 2024), peer feedback (Fan & Xu, 2020), learners' self-assessment feedback (Panadero et al., 2023), and even an integrated approach (Zhang & Hyland, 2022). A comparative study by Raheem et al. (2023) examined the relative effectiveness of these different feedback methods, providing a comprehensive overview of their impacts on student writing. Recent research (Shi & Aryadoust, 2024) has expanded to include the impact of automated written corrective feedback, enhancing our understanding of how digital tools can assist L2 learners' writing processes.

Despite its benefits, providing personalised feedback is challenging due to issues such as large class sizes and time constraints, making it a daunting task for instructors (Alharbi, 2023; Chui, 2022). To mitigate these challenges, some educators have turned to peer feedback activities, encouraging students to engage in collaborative review sessions, which were found to be beneficial in improving writing (Fan & Xu, 2020). By combining traditional feedback with automated tools, educators can offer a more comprehensive and scalable approach to improving ESL students' writing skills.

Some argue that traditional methods of instructor feedback and supervision are essential for the academic development of ESL undergraduate students. Traditional instructor feedback provides personalised attention tailored to the individual needs of students, which is crucial for addressing specific weaknesses and improving their writing skills (Sermsook et al., 2017). This type of feedback often goes beyond grammar correction to address content, organization, and critical thinking skills, which are essential for comprehensive academic growth. Moreover, direct interaction with instructors can boost students' motivation and engagement because personalised feedback helps build a rapport and creates a supportive learning environment (Hattie & Timperley, 2007).

While traditional methods of instructor feedback and supervision are undeniably beneficial for ESL undergraduate students' academic development, the integration of automated corrective feedback tools also offers significant advantages. These tools provide immediate and consistent feedback, allowing students to make timely revisions and practice self-editing skills (Shadiev & Feng, 2023; Ranalli, Link, & Chukharev-Hudilainen, 2017). Automated feedback tools support students' learning without the intensive time investment typically required for individual feedback, providing a scalable solution to address the challenges faced by educators in large class sizes.

GRAMMATICAL ERRORS IN ESL EXPOSITORY ESSAYS

Challay and Kanneh (2022) examined grammatical errors in essays by public health undergraduates, identifying significant issues in verb usage, punctuation, spelling, word choice, and capitalization. They attributed these errors to overgeneralisation, first language influence, rule ignorance, and limited English exposure, suggesting the need for early grammar exposure and positive attitudes towards English learning.

Nor Nadia Raslee et al. (2022) identified several prevalent grammatical faults in 20 ESL university students' writing. Common sentence structure concerns included issues with word order and the logical flow of ideas. Students frequently made subject-verb agreement errors related to number and person. Additionally, they struggled with using the correct tense, often mixing past, present, and future tenses. Finally, word form issues, such as using nouns instead of verbs or adjectives instead of adverbs, disrupted the grammatical accuracy of their sentences.

An investigation into the writing of 48 Malaysian diploma students by Zuraina et al. (2023) found that the highest number of errors were in tenses, subject-verb agreement, and word choices, with fewer errors in possessive structures, gerunds, and infinitives. The study implied that impromptu writing tasks and limited vocabulary knowledge contributed to these errors

Furthermore, Liong et al. (2019) conducted error analysis in ESL writing, noting frequent errors in verb tense, word order, articles, prepositions, and subject-verb agreement. They stressed the need for targeted instruction and practice to improve grammatical accuracy.

The past studies on grammatical errors in ESL writing offer valuable insights and have distinct strengths and limitations relevant to the author's study. One strength is the identification of specific error types, such as sentence structure issues, subject-verb agreement, tense inconsistencies, and word form mistakes (Nor Nadia Raslee et al., 2022; Zuraina et al., 2023; Liong et al., 2019). These studies also provide actionable recommendations for targeted instruction and practice, emphasising the importance of early grammar exposure and positive attitudes toward English learning (Challay and Kanneh, 2022). Additionally, some studies have a larger sample size, enhancing the generalisability of their findings (Zuraina et al., 2023).

However, there are notable limitations across these studies. Many have relatively small sample sizes or focus on specific student groups, limiting the generalisability of their findings to broader ESL populations (Nor Nadia Raslee et al., 2022; Challay and Kanneh, 2022). Furthermore, these studies relied on manual analysis by instructors rather than utilising automated corrective feedback tools like Grammarly and QuillBot, missing out on evaluating the potential benefits and drawbacks of these technologies. Additionally, while they identify common error types, they often do not extensively explore the underlying cognitive processes or the effectiveness of specific corrective feedback methods, which are critical for developing comprehensive ESL educational strategies.

These limitations are relevant to the present study, as it similarly faces challenges related to small sample size and focus on a specific group of students. By addressing these limitations and

extending the scope to evaluate automated corrective feedback tools, the study aims to provide more generalisable and actionable insights for improving ESL writing proficiency. This approach will help bridge the research gap by examining the efficacy of Grammarly and QuillBot in addressing common grammatical errors, thus contributing to the existing body of literature with contemporary analyses of automated grammar correction tools.

AUTOMATED GRAMMAR CORRECTIVE FEEDBACK TOOLS

Automated grammar corrective feedback (AGCF) tools, usually equipped in automated writing evaluation (AWE) systems, play a pivotal role in language learning (Shadiev & Feng, 2023). Their review of 82 articles on the use of AWCF tools revealed that Pigai, Criterion, and Grammarly were the most frequently used tools, and most studies reported positive effects of AWCF tools on language learning.

These tools are designed to aid students in honing their writing proficiency. AWE tools offer consistent, immediate feedback, manage large volumes efficiently, foster self-directed learning, and improve lexical diversity and syntactic complexity compared to human feedback (Zahra & Saman, 2023).

Furthermore, these automated tools significantly alleviate the workload on educators by eliminating the need for individualised feedback, thereby allowing them to dedicate more time to other instructional priorities (John & Woll, 2018). AWE systems facilitate a range of feedback mechanisms to the users. Raheem et al. (2023) emphasised that artificial intelligence enhances writing precision and streamlines the editing process by allowing students to correct fundamental mistakes before their work is evaluated by instructors. This enables instructors to focus on providing feedback related to content rather than basic errors. As a result, the teaching process becomes more efficient, and students receive immediate and actionable feedback on their written work.

These tools facilitate a process where students can write, receive feedback, and revise and at the same time aid teachers to reduce their time taken to meticulously analyse each of the errors identified in their students' composition (Chui, 2022). With the integration of automated corrective feedback in language classroom, students are given the opportunity to steer more on self-directed learning, provided the analysis of the errors, feedback and support by the AGCF allows them better to understand the errors they committed. Nonetheless, despite the convenience offered by the AGCFs, they are best used with human-teachers.

GRAMMARLY VS QUILLBOT

Grammarly's effectiveness as an AGCF tool is well-documented (Chen et al., 2022), alongside Pigai and Criterion, which suggests a need to explore less-studied AGCF tools (Shadiev & Feng, 2023). QuillBot, an emerging AGCF tool, has received limited attention (Chui, 2022), despite its promising potential to optimise academic writing and enhance learner competency (Raheem et al., 2023). This research gap necessitates an evaluation of QuillBot's efficacy in improving writing skills and grammatical accuracy. This study will first review Grammarly's established impact on ESL learners' writing accuracy and then explore the emerging research on QuillBot's effectiveness in the same context.

Soegiyarto et al. (2022) have shown that Grammarly plays a pivotal role in enhancing English proficiency by helping students form sentences, grasp grammar rules, and improve their writing skills. Alharbi (2023) further explains that tools like Grammarly offer real-time corrections

and suggestions as students write, primarily addressing lexical and grammatical errors. Moon (2021) concluded that Grammarly generally offers precise comments and suitable replacement suggestions, with minimal instances of incorrect alerts. Moon (2021) further deduced that the findings indicate that Grammarly has great potential as an effective educational tool to address the limitations of instructor feedback and assist learners in enhancing grammatical precision in their written assignments.

Despite Grammarly's accuracy in addressing lexical and grammatical errors, it often misses structural and organizational aspects. John and Woll (2020) highlighted Grammarly's limitations, noting its failure to identify critical errors effectively, with performance scores below 50 percent in certain areas. This low coverage means Grammarly consistently fails to detect many errors in both simple sentences and authentic ESL compositions, making it unreliable for comprehensive error detection. While proficient in detecting verbs, subject-verb agreement, plurals, and word forms, Grammarly is less reliable in evaluating complete essays, frequently marking correct elements as errors and issuing incorrect warnings more often than in simpler sentence analyses.

Nonetheless, Grammarly feedback is broadly accurate across various studies. Dodigovic and Tovmasyan (2021) observed that it generally offers accurate feedback, albeit with some inconsistencies. While the tool addresses many errors that it correctly identifies, it also misses several errors, resulting in missed opportunities for improvement. In addition to this, Vidhiasi and Haryani (2021) in their study concluded that Grammarly is still unable to quickly recognise sentences with semantic problems, which involve the meaning and appropriateness of word choices in context. This could potentially affect the overall clarity and coherence of the text. One aspect of error identification provided in Grammarly is tone modification. A study by Ambarwati (2021) concluded that participants feel Grammarly's tone modification is inaccurate and fails to comprehend the context of their texts, leading participants to disregard the input on tone modification.

Shifting the discussion to an alternative in the integration of AI in academic writing, Raheem, Anjum, and Ghafar (2023) articulated the differing emphasis of Grammarly and QuillBot. Grammarly enhances writing through improved readability feedback and tone modifications, making it ideal for creating clear, engaging developmental resources. Meanwhile, QuillBot offers a wider range of programming languages and includes features that monitor writing progression. This tool would be especially useful for individuals engaged in multilingual projects and cultivating continuous improvement in writing skills.

In addition, Chui (2022) found that the free version of QuillBot outperformed Grammarly and Ginger, another popular grammar checking tool, in accuracy on authentic student writing at both sentence and paragraph levels. This finding aligns with the research conducted by Raheem et al. (2023), which highlighted QuillBot's benefits for ESL learners, such as sentence paraphrasing, grammar and punctuation error detection, among other features made available to ESL learners.

The comparison of Grammarly and QuillBot highlights a critical need for research to assess their utility in providing corrective feedback, particularly for grammar error analysis in expository essays among ESL learners. While Grammarly has been lauded for its ability to pinpoint errors in style, lexicon, and structure with high accuracy, QuillBot introduces an alternative AGCF approach, potentially bringing its own set of advantages or challenges. This comparative study is designed to uncover which tool is more beneficial for enhancing grammatical accuracy in expository essay writing within ESL contexts. Focusing on their impacts on syntactic complexity, accuracy, and overall writing proficiency, the evaluation of Grammarly and QuillBot's

performances seeks to bridge the current knowledge gap regarding the role of AGCF tools in improving grammar in ESL expository essay

METHODOLOGY

This study employs a descriptive quantitative data analysis to gain a comprehensive understanding of the types of errors identified by both Grammarly and QuillBot in analysing the expository essays produced by five Malaysian ESL learners.

This study uses essays written by Malaysian ESL learners as its primary research instrument. The data collected is part of the students' assessment designed for their ELC 231 (Intermediate English III) subject for the third semester diploma programmes offered by the university. The subject plays a crucial role as the foundation for writing skills among diploma students. It introduces students to expository essay, a fundamental genre in academic writing, and helps them develop essential skills such as planning, writing, revising, and editing. This assessment was selected for this study because it is the first major writing assessment for their diploma level. In their sixth week of the semester, ELC231 students are tasked to write a complete five-paragraph essay, making it an ideal point for evaluating their writing abilities. Students were intensively taught basic expository essay writing skills for five weeks before this assessment, making it a good choice for evaluating automated grammar feedback tools.

In this study, a group of students from social and humanities backgrounds were asked to write an expository essay based on a given prompt; "*Write an essay of 250 - 300 words on the importance of upgrading your ability to use digital devices"* for two hours. The essays were composed by them on an online platform, known as exam.net. It was developed by a Swedish company, Teachiq AB, and designed specifically for online examinations with cheat prevention and real-time monitoring features. This platform offers three security layers to prevent cheating and plagiarism by restricting copy-paste, preventing students from using other applications during tests and alerts teachers of suspicious activity. In this study, the spell-check and dictionary functions were turned off to ensure that the participants are writing their essays without any aids.

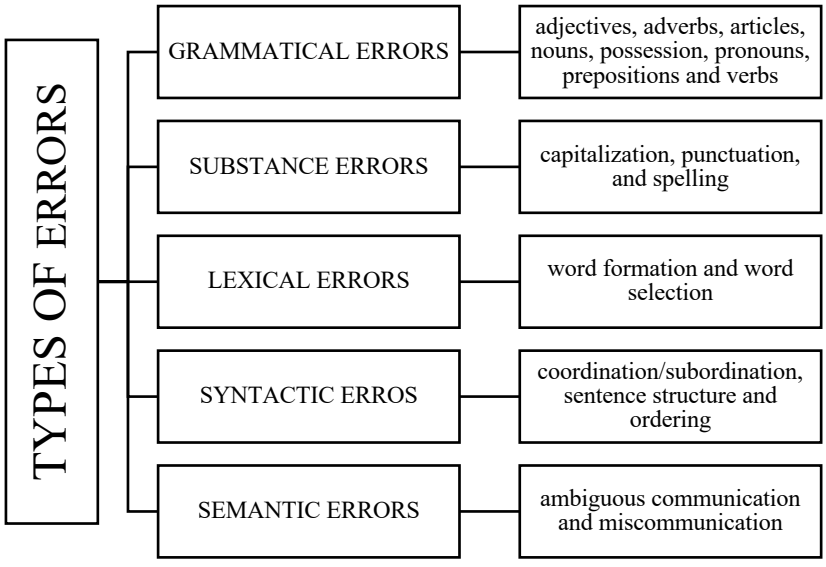
The participants in this study were from mixed levels of proficiencies, ranging from basic to independent users. The classification was based on their previous English language grade. Essays for analysis were selected using systematic sampling. Systematic sampling selects items at regular intervals from a larger population to uniformly distribute the sample across the dataset (Cochran, 1977). Firstly, 20 essays were tagged as E1–E20. Then, five essays were selected at four-essay intervals for in-depth analysis to represent the students' writing skills. Starting with essay E1, the next essays were E5, E9, E13, and E17. This approach was used to guarantee that the sample included a wide range of students' work and reduce bias from selecting essays consecutively. By doing this, clustering effects from grouping students with similar abilities or styles were avoided.

In Malaysia, the primary method for acquiring English is through its education system, which adheres to British English (BrE) standards. Despite this, Wan Noor Farah Wan Shamsuddin et al. (2019) note an interesting shift towards American English (AmE) among Malaysians, attributed to the pervasive influence of media. Nonetheless, for the purposes of this study, both platforms were configured to use British English. This decision was made to ensure uniform spelling in data analysis, thereby maintaining consistency and standardisation in the evaluation of linguistic elements.

The five selected essays were run on Grammarly and QuillBot to check for the errors. Grammarly (https://www.grammarly.com/) is an AI-powered tool, which offers a wide range of features such as grammar checker, plagiarism checker, citation generator, essay checker, tone detector, style guide, snippets, analytics and brand tones. QuillBot (https://quillbot.com/), on the other hand offers features such as paraphraser, summarizer, QuillBot flow, plagiarism checker, word counter, translator, AI detector, grammar checker, proof-reader, spell checker, punctuation checker, essay checker and citation tools.

In analysing the data, the five essays were individually uploaded into both Grammarly and QuillBot platforms and error analysis was performed. Then, the number of errors identified was recorded and tabulated at both paragraph and the essay levels into Excel sheet. Next, the types of errors detected on both platforms were analysed in detail according to the five types of errors as proposed by James (1998). James' (1998) error classification framework was used because it offers a systematic approach to analysing language errors, categorising them into five types: grammatical, substance inaccuracies, lexical, syntactic, and semantic issues. This comprehensive approach helps identify common errors and compare the effectiveness of grammar checking tools like Grammarly and QuillBot. This method enhances the reliability and validity of research findings, facilitating a clear comparison of their performance in aiding ESL learners. The classification was made according to the suggestions provided by the applications and then tabulated into a table for the total number of errors according to application used, types of errors and analysis for each type of errors. Validation of the classification of the errors was also performed by an ESL lecturer with 15 years of experience to ensure its reliability.

FIGURE 1. CLASSIFICATION OF ERRORS BY JAMES (1998)



RESULT AND DISCUSSION

THE TOTAL NUMBER OF ERRORS IDENTIFIED BY BOTH GRAMMARLY AND QUILLBOT

To answer research question one, the number of errors were identified at both whole essay and paragraph level by Grammarly and QuillBot and then tabulated into a table. The breakdown of the total number of errors identified by both applications are shown below.

TABLE 1. THE TOTAL NUMBER OF ERRORS IDENTIFIED AT THE WHOLE ESSAY LEVEL

Essay	Grammarly	QuillBot
Essay 1	111	84
Essay 5	43	46
Essay 9	67	58
Essay 13	102	94
Essay 17	65	68
Total errors	**388**	**350**

At the whole essay level, a total of 388 errors were identified by Grammarly and 350 by QuillBot. 111 errors were identified by Grammarly, while 84 errors by QuillBot in Essay 1 and 43 errors were identified by Grammarly while 46 errors by QuillBot in Essay 5. In Essay 9, 67 errors were detected by Grammarly while 58 by QuillBot. Essay 13 shows a total of 102 errors detected by Grammarly, and 94 errors by QuillBot. Finally, in Essay 17, 65 errors were identified by Grammarly, and 68 by QuillBot. A detailed analysis was done at the paragraph level to ascertain the number of error counts as detected by the two automatic grammar correction feedback platforms.

TABLE 2. THE TOTAL NUMBER OF ERRORS IDENTIFIED AT THE PARAGRAPH LEVEL

Paragraph	Grammarly	QuillBot
Paragraph 1 (Introduction)	54	65
Paragraph 2 (Body Paragraph 1)	88	74
Paragraph 3 (Body Paragraph 2)	98	83
Paragraph 4 (Body Paragraph 3)	76	59
Paragraph 5 (Conclusion)	72	68
Total errors	**383**	**359**

A total of 383 errors were identified by Grammarly and 359 errors were identified by QuillBot at the paragraph level. The detailed analysis shows that in paragraph 1 (Introduction), 54 errors were identified by Grammarly, while 65 errors by QuillBot. In paragraph 2 (Body Paragraph 1), 88 errors identified by Grammarly, while 74 errors by QuillBot. As for paragraph 3 (Body Paragraph 2), 98 errors were identified by Grammarly, and 83 errors by QuillBot. Paragraph 4 (Body Paragraph 3) indicated that Grammarly identified only 76 errors, while QuillBot identified 59 errors. While for paragraph 5 (Conclusion), a total of 72 errors were identified by Grammarly while only 68 errors were identified by QuillBot.

This finding revealed that automated grammar corrective feedback tools are very efficient in highlighting errors in students' essay. This is concurrent with Tian and Zhou (2020)'s study, which found that automated writing evaluation tools provide the most feedback in identifying the total number of errors. In this study, Grammarly identified more errors compared to QuillBot. This contrasted with the findings by Chui (2022) where QuillBot is found to outperform other grammar checkers used in the study, which are Grammarly and Ginger in detecting the number of errors. However, the study was conducted using the free version of the respective grammar checkers, thus indicating there might be some limitations with their ability in providing comprehensive error detection to the researcher. The absence of sufficient literature on the comparison between Grammarly and QuillBot set constrains to confirm the finding in this current study. Despite that, previous studies on Grammarly indicated that this tool has the ability to classify high frequency of errors in written composition (Moon, 2021; Dodigovic & Tovmasyan, 2021) and remarkable ability in identifying verbs, subject-verb agreement, plurals, and word forms (Alharbi, 2023; Tian & Zhou; 2020, Bailey & Lee; 2020 and John & Wool; 2020).

TABLE 3. COMPARISON BETWEEN TOTAL NUMBER OF ERRORS AT THE WHOLE ESSAY AND PARAGRAPH LEVEL

Levels	Grammarly	QuillBot
Whole essay	388	350
Paragraph	383	359

Table 3 shows the total number of errors detected at the whole essay and the paragraph levels for both Grammarly and QuillBot. A comparison between the applications revealed that both demonstrated inconsistencies in detecting the total number of errors at the whole essay and paragraph levels. There are variations in the number of errors when tested at the whole essay, paragraph, and sentence levels. Insights from linguistic research, such as a study on Thai EFL students' writing errors (Chuenchaichon, 2022) emphasised the nuanced capabilities of AGCF tools in contextual and grammatical analysis. For example, word-level errors like spelling are easily identified in sentences, whereas errors requiring broader context, such as coherence, are more detectable in full essays. This parallel between the study's findings and the performance of Grammarly and QuillBot underscores the complexity of automated grammar correction and the necessity of analysing text at multiple levels for thorough error detection, aligning with the need for advanced, context-aware writing assistance tools. Similarly, findings in Chui's (2022) and John and Woll's (2020) study corroborated the above finding, indicating that QuillBot exhibits inconsistent erroneous detection or numbers when tested at both the sentence and sentence levels. Likewise, Grammarly also shows inconsistent number of errors and giving incorrect warning when evaluating complete essays and simple sentences.

TYPE OF ERRORS GRAMMARLY AND QUILLBOT IDENTIFIED THE MOST IN THE ESSAYS

To answer research question two, a framework proposed by James (1998) was used to categorise the errors identified by both Grammarly and QuillBot.

TABLE 4. THE CLASSIFICATION OF ERRORS IDENTIFIED BY GRAMMARLY AND QUILLBOT

Types of error	Grammarly	QuillBot
Grammatical errors	181	161
Substance errors	76	96
Lexical errors	40	61
Syntactic errors	18	34
Semantic errors	57	7
Others	11	0
Total	383	359

Table 4 shows both applications identified mostly grammatical errors, which includes adjectives, adverbs, articles, nouns, possession, pronouns, prepositions and verbs. Grammarly identified a total of 181 errors, while QuillBot identified 161 errors under the grammatical errors. Substance errors, which include capitalisation, punctuation and spelling errors recorded the second highest number of errors identified by both applications, with a total of 76 errors by Grammarly, and 96 errors by QuillBot. This is followed by lexical errors, which include word formation and word selection with a total of 40 errors identified by Grammarly and 61 errors by QuillBot respectively. The fourth highest total number of errors identified by both applications is semantic errors, which include ambiguous communication and miscommunication with a total of 57 errors identified by Grammarly and 7 errors by QuillBot.

Finally, syntactic errors, which include coordination/subordination, sentence structure and ordering, recorded the lowest number of errors identified by both application with a total of 18 errors identified by Grammarly and 34 errors identified by QuillBot. Interestingly, there is another type of error which is 'tone' that does not belong to any types of classification by James (1998). The error was identified only by Grammarly, which was classified into others category by the researcher, with a total of 11 errors.

This finding concludes that both applications are very effective in identifying errors which are under grammatical, substance and lexical errors categories as compared to syntactic and semantic errors. This finding aligns with a study by Vidhiasi and Haryani (2021) which concluded that, despite Grammarly's proven ability to assist teachers with error analysis, the application still struggles to identify phrases with semantic mistakes. In the study, Grammarly identified the most types of errors related to spelling and punctuation. Similarly, studies by Alharbi (2023), Tian and Zhou (2020) and Bailey and Lee (2020), support this finding when they concluded that automated feedback tool provided higher feedback in surface-level feedback which focuses on writing mechanics such as articles, preposition, and verb-noun agreement more frequently, thus leaving the meaning-level feedback such as argumentation, flow, content, structural, and organizational errors untreated.

In terms of the substance, lexical and syntactic errors identified, QuillBot recorded a slightly higher number of these errors as compared to Grammarly. According to Reguig and Mouffok (2023), the developers of QuillBot created the programme in 1997 in responding to a demand for a more effective approach of composing and paraphrasing content. Hence, the primary focus of QuillBot was intentionally to assist writers to express their thoughts in their own words, while minimising the time and effort required. Chui (2022) and Raheem, et al., (2023) mention that previous studies on QuillBot were mostly focusing on its capabilities as a tool for rewriting and paraphrasing English language. Thus, this explains the fact on the insufficient literatures on QuillBot as a tool for automated grammar correction.

In this study, Grammarly is found to detect 'tone' as a type of error that should be revised in the essays. Interestingly, this error is only identified by Grammarly. Since this type of error does

not belong to any type errors as proposed by James (1998), it was classified as 'others'. Tu support this, Raheem, Anjum, and Ghafar (2023) in their study also mention that tone modification offered by Grammarly as compared to Quillbot. This finding is interesting to be noted as this feature offered by Grammarly helps the writers to convey their written ideas confidently. However, a study conducted by Ambarwati (2021) revealed that the participants think that tone modification offered by Grammarly is inaccurate and fail to detect the context of their texts as they have specific writing context to meet, hence, leaving the feedback on tone modification ignored.

FIGURE 2. A SCREENSHOT OF AN INTERFACE BY GRAMMARLY

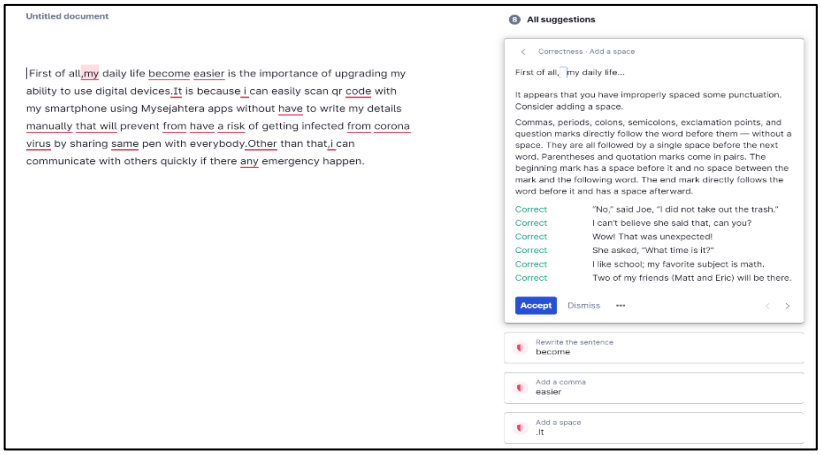


FIGURE 2. A SCREENSHOT OF AN INTERFACE BY QUILLBOT

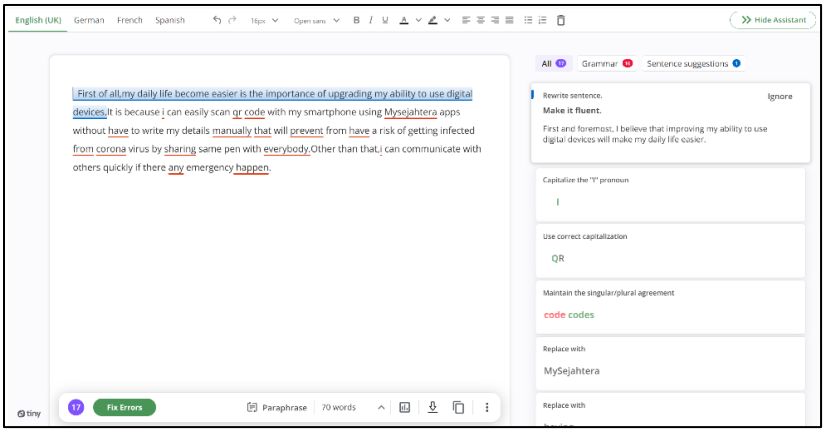


Figure 2 and Figure 3 show the interfaces of errors detected and the feedback provided by both Grammarly and Quillbot. In analysing the errors, it is interesting to note that both applications offer different interface and feedback in explaining errors to their users. To further illustrate, one paragraph of the analysed essay was chosen, and the feedback provided by both applications were compared in Table 5.

Besides, the importance of upgrading our ability to use digital devices is we can improve our knowledge. By using digital device, we can gain new knowledges via Google and any page that sharing facts and news. Therefore, digital device also can have our learning process easily which is we can learn it through websites and learning applications.

(Paragraph 3, E5)

TABLE 5. DETAILED ANALYSIS OF FEEDBACK AND CORRECTION BY GRAMMARLY AND QUILLBOT

Types of error	Grammarly	Correction	Feedback	QuillBot	Correction	Feedback
Grammatical errors	*By using digital device, we can….*	~~device~~ → devices	Incorrect noun number	*By using digital device, we can..*	*By using a digital device, we can…*	Add an article
	we can gain new knowledges via…	~~knowledges~~ → knowledge	Incorrect noun number	*we can gain new knowledges via..*	*we can gain new knowledge via..*	Use the right determiner
	..and any page that sharing facts and news.	~~sharing~~ → shares	Incorrect verb form	*..and any page that sharing facts and news.*	*..and any page that shares facts and news*	Replace with
	Therefore, digital device also can have	~~device~~ → devices	Incorrect noun number	*..our learning process easily which is we can learn it*	*..our learning process easily which ~~is~~ we can learn it*	Remove
	which is we can learn it through websites…	it	Pronoun use	*Therefore, digital device also can have*	*Therefore, digital devices can also have*	Replace with
Substance errors	*digital device also can have our learning process easily which is we can*	, which	Punctuation in compound/ complex sentences	*…digital device also can have our learning process easily which is*	*…digital device also can have our learning process, ~~easily~~ which is*	Replace with
Lexical errors	*digital device also can have our learning process*	~~also can~~ → can also	Misplaced words or phrases	*digital device also can have our learning process*	*digital device also can facilitate our learning process*	Replace with
				which is we can learn it through websites and..	*which is we can do through websites and..*	Replace with
Syntactic errors	*we can gain new knowledges via Google*	~~knowledges~~ → *understandings, acquaintances*	Word choice	*upgrading our ability to use digital devices is we can improve our knowledge*	*upgrading our ability to use digital devices is that we can improve our knowledge*	Replace with

Semantic errors	*Besides, the importance of upgrading our ability to use digital devices is we can improve our knowledge.*	*Besides upgrading our ability to use digital devices is important because we can improve our knowledge.*	Unclear sentences	*Besides, the importance of upgrading our ability to use digital devices is that we can improve our knowledge.*	*Furthermore, the significance of enhancing our proficiency with digital devices lies in its potential to enhance our knowledge base.*	Increase clarity
	By using digital device, we can gain new knowledges via Google and any page that sharing facts and news.	*Therefore, digital devices can make learning easy we can learn through websites and learning applications.*	Unclear sentences			
Others						

Table 5 depicted that both Grammarly and QuillBot offer feedback aimed at improving students' writing, focusing on grammatical, substance, lexical, syntactic, and semantic errors, as proposed by James (1998). They share a common goal of enhancing writing clarity and accuracy, but their approaches differ. For grammatical errors, Grammarly provides detailed corrections such as correcting noun numbers and verb forms, while QuillBot streamlines sentences with clear, concise suggestions. For example, Grammarly changed *"device"* to *"devices"* and add necessary articles, whereas QuillBot rephrased the same sentence to make it more fluid.

In substance errors, Grammarly's main focus is on punctuation within complex sentences, suggesting precise replacements for phrases that lack clarity. For instance, Grammarly suggested *"...digital device also can have our learning process easily which is…*" to *"...digital device also can have our learning process easily, which is we can …"*. QuillBot, on the other hand, offers comprehensive modifications for readability and coherence, by changing the sentence to *"… digital device also can have our learning process, which is..."*

Both platforms address lexical errors by correcting misplaced words, but QuillBot uses a simpler language for better understanding. For example, Grammarly suggested *"... digital device can also have our learning process "* to "*... digital device also can have our learning process…"* while QuillBot simplify this to *"… digital device also can facilitate our learning process..."*

Grammarly provides precise word selection modifications such as modifying *"...we can gain new knowledges via.."* to *"...we can gain new understandings via…"* ensuring proper word usage. QuillBot did not detect syntactic error on the same sentence, but only corrected the grammatical error there.

Lastly, Grammarly and QuillBot both help clarify semantic errors but differ in their approach, with Grammarly giving structured feedback and QuillBot simplifying complex sentences. For example, Grammarly revised *"Besides the importance of upgrading our ability to use digital devices is we can improve our knowledge"* to *"Besides upgrading our ability to use digital devices is important because we can improve our knowledge. "* QuillBot simplified this to *"Upgrading our ability to use digital devices helps us improve our knowledge."*

With regard to feedback style, Grammarly offers a detailed analysis and feedback with examples for each errors found and suggestions to its users. However, Quillbot only offers skimpy explanation and feedback on only some of the identified errors, while leaving some other errors unexplained, only with suggestion. Reguig and Mouffok (2023) mentioned that Grammarly does not only provide instantaneous feedback, highlighting errors and suggesting corrections but also identifies the problematic text, provides a detailed explanation of the issue, examples how to use them correctly and suggests fixes. The finding in this study suggests that both applications are helpful in detecting errors in the essays and is purported by Soegiyarto et al. (2022) where AI-powered learning tool is found to serve as vital in enhancing students' English skill given that it assists in helping them understand grammar rules, construct sentences, and compose them better.

CONCLUSION

This study has significant implications for education, technology, and research. Integrating tools like Grammarly and QuillBot into the curriculum provides consistent, immediate feedback, improving students' writing skills by addressing both basic and complex writing challenges. These tools are particularly valuable for ESL learners, helping them understand common errors. Educators can use them to focus on advanced writing skills. Technologically,

developers should improve the detection of semantic and syntactic errors and enhance QuillBot's user interface with better explanations. Future research should conduct comparative studies of different automated grammar correction tools to understand their strengths and weaknesses better.

The study's scientific novelty lies in its detailed comparative analysis of Grammarly and QuillBot, focusing on their effectiveness in detecting writing errors in ESL learners' expository essays. It uses James' (1998) error categorisation framework to offer a nuanced understanding of each tool's strengths and weaknesses. By examining errors at both the essay and paragraph levels, the study highlights where each tool excels or falls short. It also provides contextual insights specific to non-native English writers and identifies Grammarly's unique tone detection feature. The study contrasts its findings with previous research to highlight variability in tool performance, offering a deeper understanding of their capabilities.

This study acknowledges a few limitations, such as a small sample size of only five essays and a focus solely on expository essays by ESL learners. A larger and more diverse sample would provide more robust findings. The study recommends using tools like Grammarly and QuillBot to aid ESL learners, emphasising that these should complement, not replace, human feedback. It also suggests training teachers to integrate these tools effectively into their curriculum. Implementing these recommendations and pursuing further research will enhance the understanding of the capabilities and limitations of automated grammar correction tools.

ACKNOWLEDGEMENT

The researchers would like to express their gratitude to the participants enrolled in the ELC231 subject at Universiti Teknologi MARA Cawangan Terengganu. The researchers would also like to express their deepest thanks to the reviewers for their valuable comments on the earlier drafts of this article.

REFERENCES

Ambarwati, E. K. 2021. Indonesian university students' appropriating Grammarly for formative feedback. ELT in Focus, 3(1), 1-11, doi: 10.35706/eltinfc.v4i1.5216

Bailey, D., & Lee, A. R. 2020. An Exploratory Study of Grammarly in the Language Learning Context: An Analysis of Test-Based, Textbook-Based and Facebook Corpora. *TESOL International Journal*, *15*(2), 4-27

Challay, S. & Kanneh, V.S. 2022. Grammatical Errors in the Expository Essays of Public Health Students of Njala University. International Journal of Linguistics, 14(1):11-11. doi:10.5296/ijl.v14i1.19470

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. 2022. Two decades of artificial intelligence in education. Educational Technology & Society, 25(1), 28-47.

Chuenchaichon, Y. 2022. An Error Analysis of Written English Paragraphs at Lexical, Syntactic, and Paragraph Levels Made by Thai EFL Non-English Major Students. 3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies Vol 28(2), June 2022 http://doi.org/10.17576/3L-2022-2802-07

Chui, H. C. 2022. The QuillBot grammar checker: Friend or foe of ESL student writers. *Journal of Creative Practices in Language Learning and Teaching, 10 (1)*, 10-31.

Cochran, W. G. 1977. *Sampling Techniques* (3rd ed.). John Wiley & Sons.

Dodigovic, M., & Tovmasyan, A. 2021. Automated writing evaluation: The accuracy of Grammarly's feedback on form. *International Journal of TESOL Studies*, *3*(2), 71-87.

Fan, Y., & Xu, J. 2020. Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing, 50,* 100775. https://doi.org/10.1016/j.jslw.2020.100775

Hassan, I., Ayuni Madarina Abdul Rahman, & Mohd Nazri Latiff Azmi. 2021. Development of English Writing Skills through Blended Learning among ESL Learners in Malaysia. Arab World English Journal (AWEJ) Special Issue on CALL (7). 377-389. DOI: https://dx.doi.org/10.24093/awej/call7.26

James, C. 1998. *Errors in language learning and use: Exploring error analysis*. Routledge.

John, P., & Woll, N. 2020. Using Grammar Checkers in an ESL Context: An Investigation of Automatic Corrective Feedback. *Calico Journal*, *37*(2).

Kamaruzaman, F.M., Hamid, R., Mutalib, A.A. & Rasul, M.S. 2022. Penerokaan Spesifikasi Kemahiran Generik 4IR bagi Graduan Kejuruteraan Awam Berdasarkan Perspektif daripada Industri. *Akademika*, *Vol 92* No 3 (Isu Khas), 137–150. https://doi.org/10.17576/akad-2022-92IK1-11

Liong, Y. P. K. Abdullah., Melor Md, Yunus, Harwati Hashim. 2019. Grammatical errors in ESL writing: An error analysis. 8:302-307.

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(9). https://doi.org/10.1186/s40561-024-00295-9

Moon, D. 2021. Evaluating corrective feedback generated by an AI-powered online grammar checker. *International Journal of Internet, Broadcasting and Communication*, *13*(4), 22-29.

Nor Nadia R., Azurawati Z., Madaha H., Halimatussaadiah I., Farahidatul Akmar A., & Syaza K. 2022. Malaysian second language learners' linguistic accuracy in writing. *International Journal of Academic Research in Business and Social Sciences*, 12(10), 1351–1364. https://doi.org/10.6007/IJARBSS/v12-i10/14972

Panadero, E., García Pérez, D., Fernández Ruiz, J., Fraile, J., Sánchez-Iglesias, I., & Brown, G. T. L. (2023). University students' strategies and criteria during self-assessment: Instructor's feedback, rubrics, and year level effects. *European Journal of Psychology of Education, 38, 1031–1051. https://doi.org/10.1007/s10212-022-00639-4

Raheem, B. R., Anjum, F., & Ghafar, Z. N. 2023. Exploring the Profound Impact of Artificial Intelligence Applications (Quillbot, Grammarly and ChatGPT) on English Academic Writing: A Systematic Review. *International Journal of Integrative Research (IJIR)*, *1*(10), 599-622.

Reguig, Y., & Mouffok, A. I. 2023. *Comparative Analysis of Current Word Processing Applications That Use Artificial Intelligence Case of Study: The Use of Grammarly and Quillbot Amongst Third Year BA Students at Ibn Khaldoun University of Tiaret* (Doctoral dissertation, Université IBN KHALDOUN-Tiaret).

Sermsook, K., Liamnimitr, J., & Pochakorn, R. 2017. The Impact of Teacher Corrective Feedback on EFL Student Writers' Grammatical Improvement. *English Language Teaching*, *10*(10), 43-49.

Shadiev, R. & Feng, Y.Y. 2023. Using automated corrective feedback tools in language learning: a review study. Interactive Learning Environments, 1-29. doi: 10.1080/10494820.2022.2153145

Shi, H. & Aryadoust, V. 2024. A systematic review of AI-based automated written feedback research. *ReCALL FirstView, 1–23*. https://doi.org/10.1017/S0958344023000265

Soegiyarto, M. S., Putri, R. A., & Saputra, S. D. 2022. The Importance of Getting Automated Grammar Feedback via Grammarly, for Increasing Students' English Language Proficiency. https://doi.org/10.31219/osf.io/749a2

Su, Y., Lin, Y., & Lai, C. 2023. Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57,* 100752. https://doi.org/10.1016/j.asw.2023.100752

Tian, L., & Zhou, Y. 2020. Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System*, *91*, 102247.

Vidhiasi, D. M., & Haryani, H. 2021. The implementation of Grammarly in error analysis implementasi Grammarly dalam error analysis. *Jurnal Sains Dan Teknologi Maritim*, *21*(1), 17-25.

Wan Noor Farah Wan Shamsuddin., Ain Nadzimah Abdullah., Tan, H., & Jusan Bolong. 2019. Attitudes and driving factors behind the attitudes of Malaysians university students towards British English and American English. International Journal of Humanities, Philosophy, Language, 2 (5), 1-12.

Wondim, B. M., Bishaw, K. S., & Zeleke, Y. T. 2024. Effectiveness of teachers' direct and indirect written corrective feedback provision strategies on enhancing students' writing achievement: Ethiopian university entrants in focus. *Heliyon, 10,* e24279. https://doi.org/10.1016/j.heliyon.2024.e24279

Zhang, Z. (V.), & Hyland, K. 2022. Fostering student engagement with feedback: An integrated approach. *Assessing Writing, 51,* 100586. https://doi.org/10.1016/j.asw.2021.100586

Zuraina, A., Shahid, H. S., & Amy Zulaikha, M. A. 2023. Error analysis: Investigating the paragraph writing of ESL Malaysian learners. *Issues in Language Studies, 12*(1), 1-15. https://doi.org/10.33736/ils.5031.2023

Najlaa' Nasuha Mohd Radin (Corresponding author)
Universiti Teknologi MARA
Cawangan Terengganu, Malaysia
Email: najlaanasuha@uitm.edu.my

Aida Mustapha
Universiti Teknologi MARA
Cawangan Terengganu, Malaysia
Email: aidam254@uitm.edu.my

Aileen Farida Mohd Adam
Universiti Teknologi MARA
Cawangan Pulau Pinang, Malaysia
Email: aileen@uitm.edu.my