The Effectiveness of Using AI Tool Grammarly in Improving ESL University Students' Descriptive Writing Scores in English Proficiency Courses

Loy, P. C, Mazura Mastura Muhammad*

Department of English Language and Literature, Sultan Idris Education University, Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.1026EDU0093

Received: 08 February 2026; Accepted: 16 February 2026; Published: 25 February 2026

ABSTRACT

The current study investigates the effectiveness of using artificial intelligence (AI) tool Grammarly in improving English as a Second Language (ESL) university students' descriptive writing scores in English proficiency courses. An explanatory sequential mixed-methods design was employed, involving 150 ESL university students from five Malaysian universities across the Northern (Perak), Central (Selangor), Eastern (Terengganu), Southern (Johor), and Sarawak research zones, all of whom met the English proficiency criteria of MUET Band 3.0 or CEFR Level B1. Data were collected using a quasi-experimental comparative groups pretest–posttest design, a survey questionnaire, and semi-structured interviews. Quantitative data were analysed through independent samples t-tests, one-way ANOVA, post-hoc tests, Spearman's correlation, and descriptive analysis, while qualitative data were examined through content analysis. The quantitative findings indicate that Grammarly did not produce statistically significant improvements in overall descriptive writing scores. Nevertheless, its effectiveness varied across research zones, with the strongest impact observed in the Southern and Central zones, moderate or inconclusive effects in the Northern and Eastern zones, and negligible impact in Sarawak. A weak but statistically significant positive relationship was observed between the frequency of Grammarly usage and improvements in descriptive writing scores. Complementing these results, the qualitative data revealed that university students perceived Grammarly as a supportive tool that enhanced language accuracy, vocabulary usage, and writing confidence. Importantly, these positive perception did not consistently translate into measurable improvements in overall descriptive writing scores, particularly for higher-order writing skills, and concerns regarding over-reliance were noted. Collectively, the quantitative and qualitative findings provide a nuanced understanding of the role of Grammarly in supporting university students' development of descriptive writing scores. By addressing a research gap within the Malaysian context, the current study offers a critical, real-world perspective on the integration of Grammarly, highlighting both its pedagogical benefits and limitations. The findings provide practical insights for university students, educators, and higher Education institutions seeking to make informed decisions regarding the use of Grammarly to support the development of descriptive writing scores.

Keywords: artificial intelligence; artificial intelligence tool; ESL university students; Grammarly; descriptive writing scores

INTRODUCTION

English has emerged as the world's most widely spoken language (Solikhah et al., 2023; Naka & Spahija, 2022), with over 1.75 billion people speaking it proficiently (Angela, 2019). It is a versatile and globally acknowledged language that has evolved from humble origins to a powerful communication tool used in many fields worldwide, representing flexibility, clarity, and resilience in human communication (Manolescu, 2023). The significance of the English language extends to everyday communication, where it is the primary means of expressing ideas and concepts across linguistic boundaries (Shahzod et al., 2021).

Furthermore, English is widely used in education, with governments around the world incorporating English as a second language (ESL) or foreign language syllabus into schools and universities (Pavlyuk & Salisu, 2022). In Malaysia, English is a required course in higher education, focusing on proficiency in speaking, listening,

reading, and writing (Widianingsih & Listyaningrum, 2019). Consequently, ESL university students must master these four fundamental language skills to communicate precisely and clearly in English (Ali, 2022).

Although all four skills are important, writing is recognised as a crucial skill in language learning and is regarded as a productive skill of the English language (Mubarok et al., 2020; Ali, 2022; Anistasya, 2022). However, writing is often the most difficult skill for ESL university students to learn, as it requires thorough comprehension of the content, proper language usage, and reader-friendly syntax (Lestari, 2020). Without regular practice, university students find it challenging to write well in English (Fahmi & Cahyono, 2021), resulting in unique challenges for ESL learners.

Writing can be categorised into four styles: narrative, persuasive, expository, and descriptive (Singgih & Rachmasisca, 2020). Among these, descriptive writing is particularly important for ESL university students, as it equips them with the skill to describe a specific person, location, or item in depth (Fitri et al., 2022; Siregar, 2022; Maulidiyah & Mandarani, 2023; Wulva, 2023). It also allows students to use imagery and literary devices to cultivate a stronger connection with readers (Sharma, 2023). Despite its importance, descriptive writing is challenging due to its reliance on a wide vocabulary, precise language, and cultural context.

Artificial intelligence (AI) tools have the potential to improve ESL university students' descriptive writing scores (Younis et al., 2023). Numerous AI writing tools, such as ChatGPT, QuillBot, ProWritingAid, and Copy.ai, can enhance descriptive writing. Among them, Grammarly is the most widely known AI tool globally (Jerbi, 2023). Despite its popularity, more research is needed to determine its impact on ESL university students' descriptive writing scores.

Grammarly has the potential to significantly impact ESL students' descriptive writing scores (Abdul et al., 2022). It was specifically chosen for this study due to its advanced capabilities in improving the accuracy of descriptive text among university students (Hadiat, 2022; Faisal & Carabella, 2023). Compared to other AI tools, Grammarly effectively addresses grammatical errors, prevents plagiarism, and boosts students' writing confidence (Abdul et al., 2022). These features make Grammarly a highly relevant tool for improving ESL university students' descriptive writing scores.

What distinguishes Grammarly from other AI tools are its comprehensive error-checking features and real-time feedback (Dewi, 2023). Unlike tools with less robust error detection or delayed feedback, Grammarly provides detailed analysis and immediate corrections, which are crucial for ESL university students in overcoming descriptive writing challenges (Dewi, 2023; Armanda et al., 2022). This immediate and detailed feedback helps students produce more accurate and unambiguous texts, directly addressing the current study's aim of assessing improvements in descriptive writing scores. Therefore, Grammarly's unique features provide a compelling justification for its selection and highlight its effectiveness and relevance to the current study.

In conclusion, the four skills of English are important, with writing being the most difficult for ESL university students to master. Among the writing styles, descriptive writing is particularly important but challenging. Grammarly is an essential tool that helps university students improve their descriptive writing scores. Despite its popularity and potential strengths, research on its effectiveness in this context is limited. Hence, the current study aims to fill the research gap by investigating the effectiveness of Grammarly in improving ESL university students' descriptive writing scores in English proficiency courses.

LITERATURE REVIEW

Artificial intelligence (AI) has increasingly been applied in Education, supporting administrative tasks, teaching, and learning processes. Bibliometric analysis of publications from 2020 to 2024 using SCOPUS revealed 106 articles related to AI tools in Education. Studies indicate that AI tools enhance educational efficiency and effectiveness, providing flexible and engaging learning environments (Singh & Hiran, 2022; Osamor et al., 2023). Tools such as chatbots and AI writing assistants can assist with tasks like hypothesis formulation, summarising publications, and drafting sections of academic work, although outputs require verification to ensure accuracy and promote critical thinking skills.

A focused analysis of Malaysian studies revealed that research on AI tools in the local context is limited. Only two relevant studies were identified. Rahim et al. (2022) examined factors influencing AI chatbot adoption in Malaysian higher education, finding that interaction, design, and ethics shape perceived trust, while behavioural intention is influenced by trust, performance expectancy, and habitual use. Dahri et al. (2024) investigated the impact of AI tool adoption on student satisfaction and academic performance in Malaysia and Pakistan, reporting positive effects on learning outcomes and engagement. These studies indicate the potential of AI tools but highlight the scarcity of research within Malaysia.

ESL university students face specific challenges in descriptive writing, which AI tools can potentially address. A bibliometric search across SCOPUS, ERIC, and SpringerLink (2020–2024) examining Grammarly's impact on ESL students' descriptive writing identified 14 articles. However, the majority focused on ChatGPT, with no studies specifically evaluating the effectiveness of Grammarly, particularly in Malaysia. This highlights a significant research gap in the use of Grammarly for enhancing ESL students' descriptive writing skills.

In summary, while global research demonstrates that AI tools can support teaching and learning, Malaysian studies are limited, and there is no evidence of research on Grammarly's role in improving descriptive writing scores among ESL university students. The current study aims to address this gap by investigating the effectiveness of Grammarly in enhancing ESL students' descriptive writing scores in English proficiency courses.

METHODOLOGY

The current study adopted an explanatory sequential mixed-methods design. In the quantitative phase, a quasi-experimental comparative-groups pretest–posttest design was employed, followed by a qualitative phase aimed at elaborating and contextualising the quantitative findings.

The current study involved 150 ESL university students from five Malaysian universities across the Northern (Perak), Central (Selangor), Eastern (Terengganu), Southern (Johor), and Sarawak zones. All university students met the English proficiency criteria of MUET Band 3.0 or CEFR Level B1. They were assigned to either experimental groups, which used Grammarly, or control groups, which did not use the AI tool.

Data for the current study were collected using multiple instruments. A descriptive writing test aligned with MUET and CEFR standards was administered to measure students' writing performance. Additionally, a survey questionnaire captured university students' perception of Grammarly, complemented by semi-structured interviews that provided deeper insight into its effectiveness in improving descriptive writing skills.

Quantitative data in the current study were analysed using independent samples t-tests, one-way ANOVA, post-hoc tests, Spearman's correlation, and descriptive analysis. Qualitative data from interviews were analysed using content analysis to identify recurring themes and provide deeper context to the quantitative findings.

Descriptive Writing Test

The descriptive writing test was the primary instrument in this quasi-experimental study, administered to both experimental and control groups during the pretest and posttest phases. University students were required to write a 250-word essay on the topic *"My Experience in the University"* within 50 minutes. The task was aligned with MUET Band 3.0 or CEFR B1 standards (Malaysian Examinations Council, 2019; Council of Europe, 2018). Pretest and posttest conditions, instructions, and timing were standardized across all university students to ensure fairness and control for testing variability.

Essays were evaluated using a validated analytic rubric adapted from Brown (2007), covering three dimensions: Vocabulary, Mechanics, and Organization (see Table 1). The rubric was reviewed by two ESL experts with over five years of teaching experience to ensure content and construct validity. Each dimension was scored on a 1–4 scale, giving a total score range of 3–12, with higher scores indicating stronger descriptive writing performance.

Two expert raters independently scored all essays using the rubric. Inter-rater reliability was assessed using the intraclass correlation coefficient (ICC = 0.87), indicating strong agreement between raters. Any discrepancies

were resolved through discussion to reach consensus, ensuring consistent application of the rubric across pretest and posttest phases.

To complement human scoring, ChatGPT was employed to automatically process essays according to the same rubric. The ICC analysis demonstrated substantial agreement between human raters and ChatGPT. Regular cross-checking helped mitigate potential threats, such as rater fatigue or differences between human and AI scoring, thus ensuring reliable and valid measurement of descriptive writing performance.

Overall, this rigorous scoring procedure ensured that any observed changes in descriptive writing scores could be attributed to the intervention. It also provided a solid basis for subsequent statistical analysis, including t-tests, ANOVA, effect sizes, and correlation assessments.

Table 1. Descriptive Writing Scoring Rubric Developed by Brown (2007)

No.	Aspects	Scores	Criteria
1	Vocabulary (V)	4	Effective choice of words and word form.
		3	Few misuses of vocabularies and words, but not change the meaning.
		2	Limited range of confusing words and word form.
		1	Very poor knowledge of words, word form, and not understandable.
2	Mechanics (M) Punctuation Capitalization	4	It uses correct punctuation and capitalization.
		3	It has occasional errors in punctuation and capitalization.
		2	It has frequent errors in punctuation and capitalization.
		1	It is dominated by errors in punctuation and capitalization.
3	Organization (O) Identification Description	4	Identification is complete, and descriptions are arranged with proper connectives.
		3	Identification is almost complete, and descriptions are arranged with almost proper connectives.
		2	Identification is not complete, and descriptions are arranged with a few misuses of connectives.
		1	Identification is not complete, and descriptions are arranged with a misuse of connectives.

Survey Questionnaire

A quantitative survey questionnaire was employed to investigate ESL university students' perception of Grammarly regarding its effectiveness in improving descriptive writing skills. The questionnaire, administered via Google Forms, consisted of fifteen closed-ended items using a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). It was structured into two sections. Section A collected demographic information, including age, gender, native language, prior experience with Grammarly, and use of other AI tools for descriptive writing. Section B assessed university students' perception of Grammarly's effectiveness in supporting grammar, vocabulary, creativity, writing confidence, and performance in descriptive writing.

Content validity was established through expert review by two English language educators with over five years of teaching experience. Each item was evaluated for relevance, clarity, coverage, and specificity using a four-

point scale (1 = poor to 4 = excellent), ensuring that the instrument accurately measured the intended constructs. Reliability was verified through a pilot test involving ten university students from the target population. The overall Cronbach's alpha coefficient was 0.803, indicating good internal consistency and confirming that the questionnaire was a reliable instrument for the current study.

Semi-Structured Interviews

To obtain in-depth insights into ESL university students' experiences with Grammarly, semi-structured interviews were conducted. The interview questions were developed based on survey responses and focused on grammatical support, feedback utility, vocabulary development, creative expression, and willingness to recommend Grammarly to peers. All interviews were audio-recorded, transcribed verbatim, and analysed qualitatively to capture university students' perspectives accurately.

Content validity was confirmed by the same two English language experts who reviewed the survey questionnaire. Each question was assessed for relevance, clarity, coverage, and specificity using a four-point scale, ensuring that the interviews appropriately measured the intended constructs. Reliability was strengthened through triangulation, by comparing interview data with survey responses, and member checking, in which university students reviewed their transcriptions to confirm accuracy and authenticity. These procedures ensured that the qualitative data were both credible and trustworthy.

External Threats to Validity

The current study identified several potential external threats to validity, including history, maturation, instrumentation, testing, mortality, selection bias, regression to the mean, subject effects, and experimenter effects. Historical events during the current study period could have influenced the university students' descriptive writing scores; to mitigate this, the intervention was kept brief (three weeks), and a control group that did not use Grammarly was included to account for external factors. Maturation effects were controlled by comparing pretest and posttest scores within the same timeframe, ensuring that natural developmental changes were distinguished from intervention effects. Instrumentation and testing threats were addressed through the consistent use of a standardized descriptive writing test and analytic scoring rubric, while assessments were administered under uniform conditions with clear instructions. To reduce the impact of university student dropout, the current study ensured clear communication of requirements and provided participation incentives. Random assignment of university students to experimental and control groups prevented selection bias, and statistical techniques accounted for regression to the mean among extreme pretest scores. Potential subject effects were minimized by blinding university students to the specific hypotheses of the current study, and experimenter effects were controlled through standardized procedures and inter-rater reliability checks. Collectively, these measures strengthened the current study's internal and external validity, allowing observed improvements in descriptive writing scores to be confidently attributed to the use of Grammarly.

RESULTS AND FINDINGS

This section discusses the quantitative and qualitative findings of the current study.

RQ1: Is there a statistically significant difference in the improvement of descriptive writing scores between ESL university students who use Grammarly (experimental group) and those who do not use Grammarly (control group) within each Malaysian university?

RQ1 examined whether the use of Grammarly led to statistically significant improvements in descriptive writing scores across five Malaysian research zones. Independent samples t-tests were conducted for each zone. Descriptive writing was assessed using Brown's (2007) analytic rubric, with inter-rater reliability of Cohen's $\kappa = 0.85$, ensuring consistent scoring. It is acknowledged that contextual factors, such as teaching quality, curriculum emphasis, and university students' digital literacy, may have influenced the outcomes.

In the Northern zone (Perak), both the control and experimental groups demonstrated improvement from pretest to posttest. The posttest difference between groups was not statistically significant ($t(28) = 0.35$, $p = .733$,

Cohen's d = 0.09), suggesting that Grammarly did not lead to greater improvement in descriptive writing scores in this zone.

In the Central zone (Selangor), a statistically significant difference emerged in posttest scores; however, the control group outperformed the experimental group ($t(19.02) = 3.174$, $p = .005$, Cohen's $d = 1.42$). This indicates that Grammarly use did not enhance descriptive writing scores here, and other factors may have contributed to the control group's superior gains.

In the Eastern zone (Terengganu), the experimental group started with significantly higher pretest scores ($t(22.84) = -3.03$, $p = .006$, Cohen's $d = 1.16$). Posttest scores revealed no statistically significant difference between groups ($t(18.03) = -1.87$, $p = .078$, Cohen's $d = 0.44$), indicating that although the experimental group began with stronger baseline skills, Grammarly did not result in a significant improvement over the control group.

In the Southern zone (Johor), while the control group consistently achieved higher mean scores in both pretest and posttest, differences were not statistically significant ($t(28) = 1.57$, $p = .127$, Cohen's $d = 0.45$). These findings suggest that Grammarly did not contribute to measurable gains in descriptive writing scores in this zone.

In the SQL zone (Sarawak), the control group outperformed the experimental group at pretest ($t(24.27) = 3.71$, $p = .001$, Cohen's $d = 1.49$), but by posttest, the experimental group had slightly higher mean scores. This posttest difference was not statistically significant ($t(28) = -0.37$, $p = .715$, Cohen's $d = 0.11$), indicating negligible effect of Grammarly use in improving descriptive writing scores in this zone.

Across all five zones, Grammarly did not consistently lead to statistically significant improvements in descriptive writing scores. While some experimental groups showed numerical gains, these differences were not significant, and in the Central zone, the control group significantly outperformed the experimental group. These results highlight the potential influence of contextual factors, such as instructional practices, university student motivation, and curriculum design, as well as the limitations of Grammarly in supporting higher-order aspects of descriptive writing, including idea generation, elaboration, and creative expression.

RQ2: How does the effectiveness of Grammarly in improving descriptive writing scores vary across different research zones in Malaysia, based on comparisons between the experimental and control groups?

RQ2 examined how the effectiveness of Grammarly varied across Malaysian research zones, based on posttest score comparisons between the experimental and control groups. One-way ANOVA followed by post-hoc analysis were conducted to identify zone-based differences in mean gains in descriptive writing scores. Effect sizes were calculated using Cohen's d, and all assessments relied on Brown's (2007) analytic rubric with inter-rater reliability of $\kappa = 0.85$. Contextual variables such as teaching quality, curriculum emphasis, and university students' digital literacy were considered as potential influences on observed differences.

In the Northern zone (Perak), the experimental group recorded a small mean gain in descriptive writing scores ($M = 2.50$, 95% CI [−3.44, 8.44], Cohen's $d = 0.18$), which was not statistically significant. This suggests that Grammarly had minimal impact on descriptive writing score improvement in this zone.

In the Central zone (Selangor), despite some numerical gains in the experimental group ($M = 11.94$, 95% CI [6.00, 17.87], Cohen's $d = 0.88$), posttest results revealed that the control group outperformed the experimental group. While the effect size indicates a moderate improvement for the experimental group, the statistically significant advantage of the control group underscores that Grammarly did not enhance performance relative to non-users in this zone.

In the Eastern zone (Terengganu), the experimental group showed modest gains ($M = 3.59$, 95% CI [−2.35, 9.52], Cohen's $d = 0.25$), but the confidence interval included zero, indicating statistical inconclusiveness. The effectiveness of Grammarly in this zone appears limited and inconsistent.

In the Southern zone (Johor), the experimental group in this zone achieved the highest mean gain ($M = 15.28$,

95% CI [9.35, 21.22], Cohen's d = 1.05), representing a large effect size. This finding suggests that Grammarly contributed meaningfully to improvements in descriptive writing score in the Southern zone, with statistical evidence supporting its positive impact.

In the SQL zone (Sarawak), a slight negative mean gain was observed in the experimental group (M = −2.50, 95% CI [−8.44, 3.43], Cohen's d = −0.16), indicating that Grammarly was ineffective in improving descriptive writing scores in this zone. No statistically significant improvement was detected.

The results indicate that Grammarly's effectiveness varies across research zones. Its impact was strongest in the Southern zone (Johor), moderate but inconclusive in the Central zone (Selangor), minimal in the Northern (Perak) and Eastern (Terengganu) zones, and negligible in the SQL zone (Sarawak). These variations highlight the influence of contextual factors such as teaching practices, student engagement, baseline writing ability, and institutional support. While Grammarly may provide benefits in certain contexts, its effectiveness is not uniform across Malaysian universities.

RQ3: Is there a relationship between the frequency of Grammarly usage and the improvement in descriptive writing scores among ESL university students who use Grammarly in the experimental group?

RQ3 examined whether the frequency of Grammarly use among ESL university students in the experimental group (N = 75) was associated with improvements in descriptive writing scores. Spearman's correlation analysis was conducted to assess the relationship between the frequency of Grammarly usage and the improvement in descriptive writing scores.

The analysis revealed a weak but statistically significant positive correlation (r_s = 0.248, p = .032), indicating that university students who used Grammarly more frequently tended to achieve slightly higher improvements in their descriptive writing scores. The observed correlation strength corresponds to a small effect size, suggesting that while usage frequency has a measurable impact, it accounts for only a limited portion of the variance in descriptive writing score improvements.

These findings imply that regular engagement with Grammarly can modestly enhance descriptive writing outcomes; however, additional factors—such as baseline writing proficiency, motivation, and instructional support—likely play a more substantial role in university students' overall improvement. Consequently, Grammarly usage should be considered a complementary tool rather than a sole determinant of writing success.

University students who actively integrate Grammarly into their writing process may benefit from incremental gains in grammar, vocabulary, and stylistic accuracy, but meaningful improvements in descriptive writing require holistic engagement with content development, idea generation, and higher-order writing skills.

RQ4: What is the perception of ESL university students across Malaysian universities who use Grammarly regarding its effectiveness in improving their descriptive writing skills in English proficiency courses?

RQ4 explored ESL university students' perception of Grammarly's effectiveness in enhancing their descriptive writing skills. Data were collected from 75 university students in the experimental group via survey questionnaires, while 15 volunteer university students participated in semi-structured interviews. Both descriptive and content analysis were conducted to provide a comprehensive understanding of university students' experiences with Grammarly.

The descriptive analysis of survey questionnaire data indicated a generally positive perception of Grammarly. Most university students were aged 21 to 23, with females representing 72% of respondents, and Malay as the predominant native language (89.3%). Prior experience with Grammarly was high (88%), and among alternative AI tools, ChatGPT was most frequently used, followed by QuillBot. University students reported that Grammarly improved their descriptive writing skills by enhancing grammar, vocabulary, clarity, and confidence. They indicated that Grammarly's feedback was timely, clear, and helpful in correcting errors, generating more engaging descriptions, saving time, and supporting better performance in descriptive writing tasks. Almost all

university students expressed willingness to recommend Grammarly to peers for improving descriptive writing skills.

Content analysis of interview data revealed two overarching themes: the strengths and weaknesses of Grammarly. University students highlighted that Grammarly improved grammatical accuracy, particularly in subject–verb agreement and tense usage, and enhanced vocabulary richness through synonym suggestions. Grammarly was also credited with refining stylistic appropriateness, promoting conciseness, readability, and a more formal tone. University students described the tool as fostering self-correction and independent editing, helping them avoid repeated mistakes, while also increasing their confidence and motivation in descriptive writing, supported by improved scores and reduced anxiety about mechanics.

Despite these strengths, university students also identified several limitations. Occasional inaccurate suggestions could alter the intended meaning, highlighting the need for critical evaluation of Grammarly's feedback. Over-reliance on the tool was another concern, as some university students reported reduced attention to independent checking. Accessibility constraints were noted, with the premium version considered costly and the free version limited in features. In terms of style, some university students reported that Grammarly sometimes made their writing too formal, rigid, or robotic, limiting creative expression. Finally, university students observed that Grammarly offered limited support for idea generation and content development, leaving the responsibility for creative and structural aspects of writing on the university students themselves.

In conclusion, ESL university students widely perceived Grammarly as an effective tool for improving grammar, vocabulary, clarity, and confidence in descriptive writing. At the same time, its limitations—including occasional inaccuracies, dependency risks, financial barriers, stylistic rigidity, and lack of creative input—underscore the importance of using Grammarly critically and as a supplement to independent editing rather than as a replacement for higher-order writing skills.

DISCUSSION

This section discusses the findings of the current study, commencing with the first research question (RQ1).

RQ1: Is there a statistically significant difference in the improvement of descriptive writing scores between ESL university students who use Grammarly (experimental group) and those who do not use Grammarly (control group) within each Malaysian university?

The findings for RQ1 indicate that Grammarly did not consistently produce statistically significant improvements in descriptive writing scores across the Malaysian universities in the current study. In the Northern, Eastern, Southern, and SQL (Sarawak) zones, differences between experimental and control groups were not significant, although modest improvements were observed. In the Central zone, a statistically significant difference favoured the control group, highlighting the context-dependent nature of descriptive writing outcomes. Factors such as instructional practices, student motivation, and emphasis on content development may have influenced performance beyond Grammarly use.

Descriptive writing requires idea development, sensory detail, coherent organisation, and stylistic expression in addition to grammatical accuracy. Grammarly primarily supports surface-level mechanics such as grammar, spelling, and sentence structure, offering limited guidance on higher-order processes like idea generation and creative expression. Consequently, improvements in mechanics may not directly translate into overall descriptive writing score gains.

These findings align with Vygotsky's Zone of Proximal Development, where Grammarly acts as a "more knowledgeable other," providing corrective feedback and scaffolding. However, its support is insufficient for higher-order descriptive writing skills, indicating that AI tools should supplement, not replace, guided instruction. Overall, the results suggest that Grammarly is most effective as a tool for improving mechanical aspects of writing, while meaningful score gains in descriptive writing require combined support from instructors and content-focused strategies.

RQ2: How does the effectiveness of Grammarly in improving descriptive writing scores vary across different research zones in Malaysia, based on comparisons between the experimental and control groups?

The findings for RQ2 indicate that the effectiveness of Grammarly in improving descriptive writing scores varied across Malaysian research zones. Grammarly had the strongest impact in the Southern and Central zones,

moderate or inconclusive gains in the Northern and Eastern zones, and minimal effect in the SQL (Sarawak) zone. These variations highlight that the benefits of Grammarly are context-dependent and shaped by regional factors such as institutional support, access to technology, and the extent of university students' engagement with the tool.

The observed differences can be explained by geographically rooted contextual factors. Southern and Central zones, home to more urbanised institutions with stable internet, stronger technological infrastructure, and greater availability of digital learning initiatives, allowed university students more consistent access to Grammarly and opportunities to engage in technology-supported descriptive writing practice. In contrast, Northern, Eastern, and SQL zones faced limitations due to semi-urban or rural settings, uneven digital literacy, and restricted access to technology-enhanced learning resources, which constrained university students' use of Grammarly and limited score improvements.

These findings align with Bandura's Social Cognitive Theory (SCT), which emphasises self-efficacy, observational learning, and reciprocal determinism. University students in zones with stronger digital literacy and more exposure to technology-supported environments likely had higher technological self-efficacy, enabling them to use Grammarly more effectively. Conversely, university students in zones with limited access and lower digital confidence may have engaged with Grammarly less strategically, reducing its impact on their descriptive writing scores. SCT further highlights the importance of environmental factors, showing that institutional support, enrichment programmes, and technology-integrated instruction contribute to the uneven effectiveness observed across zones.

The context-dependent variation observed in RQ2 complements the RQ1 findings. While RQ1 revealed mostly non-significant differences between experimental and control groups within individual zones, RQ2 demonstrates that Grammarly's relative effectiveness differs across regions, emphasising that its impact is shaped by broader instructional and environmental factors rather than uniform performance gains.

Overall, the current study demonstrates that Grammarly's effectiveness in improving descriptive writing scores is influenced by a combination of personal, behavioural, and contextual factors. University students with higher self-efficacy and access to supportive, technology-rich environments benefited more from the tool, whereas regional disparities limited its impact in other zones. These results underscore the importance of combining AI tools with guided instruction and targeted institutional initiatives to optimise their effectiveness, particularly in multi-regional ESL contexts.

RQ3: Is there a relationship between the frequency of Grammarly usage and the improvement in descriptive writing scores among ESL university students who use Grammarly in the experimental group?

The findings for RQ3 reveal a weak but statistically significant positive relationship between the frequency of Grammarly usage and improvement in descriptive writing scores among ESL university students (r_s = .248, p = .032). This indicates that university students who used Grammarly more frequently tended to achieve slightly greater gains in their descriptive writing performance. However, the modest strength of the correlation suggests that frequency of use alone is insufficient to produce substantial improvements in overall descriptive writing scores.

This outcome reflects the functional scope of Grammarly as a form-focused writing tool. Grammarly provides immediate feedback on grammar, spelling, and sentence construction, which supports surface-level accuracy but offers limited guidance on higher-order writing skills such as idea development, organisation, and stylistic expression. As descriptive writing is a content-heavy genre requiring both linguistic accuracy and rich elaboration of ideas, improvements in mechanical accuracy do not necessarily translate into marked gains in overall writing scores, even with frequent use.

From a theoretical perspective, these findings align with Bandura's Social Cognitive Theory (SCT), which emphasises the role of repeated practice, self-regulation, and observational learning in skill development. Frequent exposure to corrective feedback may enable university students to recognise patterns and internalise grammatical conventions. At the same time, the results also reflect the Technology Acceptance Model (TAM),

as university students who perceived Grammarly as useful and easy to use were more likely to engage with it consistently, leading to incremental learning gains.

Nevertheless, the weak correlation also highlights the potential risk of overreliance on automated feedback. University students who engage with Grammarly passively, without critically reflecting on revisions or developing independent writing strategies, may experience limited progress beyond surface-level improvements. This suggests that the effectiveness of Grammarly is contingent on active, reflective, and purposeful engagement, supported by appropriate instructional guidance.

Overall, the findings for RQ3 indicate that while frequent Grammarly use contributes to measurable improvements in descriptive writing scores, its impact remains limited without complementary pedagogical support. Grammarly functions best as a supportive scaffold rather than a standalone solution, reinforcing the need for guided instruction to foster higher-order descriptive writing skills among ESL university students.

RQ4: What is the perception of ESL university students across Malaysian universities who use Grammarly regarding its effectiveness in improving their descriptive writing skills in English proficiency courses?

The integrated quantitative and qualitative findings for RQ4 reveal a clear divergence between ESL university students' positive perception of Grammarly and the absence of statistically significant improvements in their descriptive writing scores. Survey and interview data indicate that university students perceived Grammarly as highly effective in supporting grammar accuracy, vocabulary enrichment, writing confidence, fluency, and clarity. However, these favourable perceptions did not translate into measurable gains in descriptive writing scores within the duration of the intervention.

This divergence highlights an important distinction between perceived effectiveness and objective performance outcomes. University students' positive perception appear to reflect improvements in the writing process—such as increased confidence, greater awareness of language accuracy, and enhanced self-correction—rather than substantive mastery of descriptive writing skills. Grammarly enabled university students to focus less on mechanical errors and more on idea expression; nevertheless, these process-oriented benefits did not result in significant score improvements, underscoring that subjective satisfaction does not necessarily equate to measurable performance gains.

From a pedagogical perspective, Grammarly functions primarily as a scaffolding tool rather than a direct performance enhancer. This aligns with Vygotsky's Zone of Proximal Development (ZPD), as Grammarly provided real-time support that facilitated independent revision within university students' learning zones. Similarly, Social Cognitive Theory (SCT) explains how Grammarly's immediate feedback strengthened self-efficacy and motivation, while the Technology Acceptance Model (TAM) accounts for the high level of acceptance driven by perceived usefulness and ease of use. However, while these theoretical frameworks explain positive perception and engagement, they do not guarantee improvements in higher-order writing outcomes.

The findings further suggest that descriptive writing—being a content-heavy and cognitively demanding genre—requires more than linguistic accuracy alone. Grammarly offers limited support for idea development, organisation, and stylistic creativity, which are central to high-quality descriptive writing. Overreliance on automated feedback, without active reflection or instructional guidance, may therefore reinforce surface-level improvements while constraining deeper cognitive engagement.

Overall, the findings for RQ4 indicate that ESL university students perceive Grammarly as a valuable supplementary learning tool that enhances confidence, accuracy, and writing awareness. However, positive perception alone are insufficient to ensure measurable improvements in descriptive writing scores. These findings emphasise the need for balanced integration of Grammarly with explicit instruction, guided practice, and higher-order writing support to foster meaningful development in ESL university students' descriptive writing.

Limitations Of The Study

The current study acknowledges several limitations that may have influenced its scope, findings, and generalisability.

First, the institutional scope was limited to five Malaysian universities across the Northern (Perak), Central (Selangor), Eastern (Terengganu), Southern (Johor), and Sarawak zones. Although this allowed comparative analysis across zones, it did not capture the full diversity of Malaysian higher Education institutions. Differences in curriculum design, teaching practices, and technological readiness may limit generalisation to other universities.

Second, the sample size of 150 ESL university students, while moderate, may not fully represent the broader population of Malaysian ESL university students. A larger and more diverse sample would enhance external validity and strengthen generalisability.

Third, the qualitative component included only 15 voluntary university students from the experimental group. Although sufficient to identify recurring themes, this number may not reflect the full range of university students' perception regarding Grammarly's effectiveness in improving descriptive writing skills.

Fourth, the current study examined only one AI tool, Grammarly. While this allowed an in-depth investigation, the findings cannot be generalised to other AI tools, such as ChatGPT, QuillBot, or ProWritingAid, which may offer different features or pedagogical benefits.

Fifth, the current study focused exclusively on descriptive writing. This limitation restricts generalisability to other writing genres, such as argumentative, expository, or academic essays. Future research could include multiple genres to assess the broader applicability of AI-assisted writing interventions.

Sixth, the current study duration was relatively short, making it difficult to evaluate long-term effects of Grammarly on descriptive writing scores. Longer-term studies would provide insight into sustained improvement and tool efficacy over time.

Seventh, the current study did not incorporate explicit metacognitive training alongside Grammarly use. As a result, university students' development of higher-order writing skills—such as planning, monitoring, and evaluating their own writing—may have been limited, and over-reliance on the AI tool could have occurred. Future research could investigate the combined effect of AI tools and metacognitive strategies on writing performance.

Finally, ethical considerations, including data privacy and security, were a concern due to the use of AI tools. Risks of data loss, unauthorized access, or misuse highlight the need for stringent safeguards in future research.

Despite these limitations, the current study provides valuable insights into the effectiveness of Grammarly in enhancing ESL university students' descriptive writing scores and serves as a foundation for future investigations.

Suggestions For Future Research

In light of the limitations identified in the current study, several suggestions are proposed to strengthen understanding of how AI tools—particularly Grammarly—can enhance ESL university students' descriptive writing scores in English proficiency courses.

First, future studies should broaden institutional coverage by including more public and private universities across Malaysia. While the current study focused on five public universities representing the Northern (Perak), Central (Selangor), Eastern (Terengganu), Southern (Johor), and Sarawak zones, this did not capture the full diversity of Malaysian higher Education institutions. Expanding institutional representation would improve

generalisability and provide a more comprehensive view of how contextual factors, such as curriculum design, teaching practices, and technological readiness, influence the effectiveness of Grammarly.

Second, increasing the sample size and diversity of university students is recommended. A larger and more heterogeneous sample, including students of varying English proficiency levels, academic backgrounds, and demographics, would enhance external validity and offer a more detailed understanding of how Grammarly supports descriptive writing improvements across different learner groups.

Third, future research should expand the qualitative component to capture a wider range of perspectives. In the current study, only 15 volunteered university students from the experimental group participated in semi-structured interviews. Including more university students and employing additional qualitative methods—such as focus group discussions or classroom observations—would provide deeper insights into university students' attitudes, learning strategies, and challenges when using Grammarly. Such triangulation would strengthen the depth and credibility of qualitative findings.

Fourth, researchers are encouraged to explore and compare multiple AI writing tools beyond Grammarly. Tools such as ChatGPT, QuillBot, and ProWritingAid offer unique features that may differently support descriptive writing development. Comparative studies could identify which AI tools or combinations are most effective in improving ESL university students' descriptive writing scores.

Fifth, longitudinal research designs should be employed to examine the long-term effects of Grammarly on descriptive writing development. The relatively short duration of the current study limited the ability to assess sustained improvement and long-term writing autonomy. Longer-term studies would allow researchers to determine whether continued exposure to Grammarly leads to permanent gains in descriptive writing scores and greater self-regulation in writing.

Sixth, integrating metacognitive training alongside AI tool use is recommended. Teaching university students to plan, monitor, and evaluate their own writing while using AI tools such as Grammarly or ChatGPT can reduce over-reliance on automated feedback and support the development of higher-order writing skills, including critical thinking, organization, and self-regulation. Future studies could investigate the combined effect of AI tools and metacognitive strategies on descriptive writing performance.

Finally, future studies should address ethical considerations in AI tool usage. Given that Grammarly processes and stores university students' data online, potential concerns related to privacy, storage, and security must be addressed. Comprehensive ethical protocols and institutional guidelines should be implemented to ensure data confidentiality and responsible use of AI in educational contexts.

In conclusion, future research should aim to broaden institutional representation, increase university student diversity, strengthen qualitative inquiry, compare multiple AI tools, examine long-term effects, integrate metacognitive training, and enhance ethical safeguards. Addressing these areas will allow subsequent studies to build on the current research, providing a more robust, generalisable, and comprehensive understanding of how AI tools can effectively improve ESL university students' descriptive writing scores in English proficiency courses.

CONCLUSION

The current study provides empirical insight into the effectiveness of the AI tool Grammarly in improving ESL university students' descriptive writing scores within English proficiency courses. Addressing a clear research gap within the Malaysian context, the current study offers a critical, real-world perspective on the integration of Grammarly, highlighting both its pedagogical strengths and weaknesses. The findings also provide practical insights for university students, educators, and higher Education institutions seeking to make informed decisions regarding the use of Grammarly to support the development of descriptive writing scores.

The results indicate that while Grammarly functioned as a useful supplementary writing tool, it did not produce statistically significant improvements in descriptive writing scores between experimental and control groups

across the participating Malaysian universities. The effectiveness of Grammarly also varied across research zones, with university students in the Southern and Central zones showing relatively greater improvement than those in other zones, suggesting that contextual factors such as institutional support, digital literacy, and learning environments may influence its impact. In addition, a weak but statistically significant positive relationship was identified between the frequency of Grammarly usage and improvement in descriptive writing scores, indicating that consistent engagement with the tool contributes to gradual learning gains, albeit to a limited extent.

Despite these modest quantitative outcomes, university students reported overwhelmingly positive perception of Grammarly. They perceived the tool as effective in enhancing grammatical accuracy, vocabulary use, writing confidence, and awareness of language mechanics. However, university students also acknowledged several limitations, including occasional inaccurate feedback, overreliance on automated corrections, restricted access to premium features, stylistic rigidity, and limited support for creativity and higher-order writing skills.

Overall, the findings suggest that Grammarly is best positioned as a supportive pedagogical resource rather than a replacement for instructor feedback or higher-order writing instruction. While the tool can facilitate accuracy, confidence, and autonomous learning, meaningful improvement in descriptive writing requires guided instruction, critical engagement, and sustained practice. Future research is recommended to examine the long-term impact of Grammarly use and to explore pedagogical models that integrate AI-based feedback with human instruction to support the holistic development of university students' descriptive writing scores.

ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Abdul et al. (2022). Exploring artificial intelligence using Automated Writing Evaluation (AWE) for writing skills. Environment-Behaviour Proceedings Journal, 7(SI9), 547–553. https://doi.org/10.21834/ebpj.v7isi9.4304
2. Ali, H. (2022). The importance of the four English language skills: Reading, writing, speaking, and listening in teaching Iraqi learners. Humanitarian and Natural Sciences Journal, 3(2). https://doi.org/10.53796/hnsj3210
3. Angela, C. (2019). English the lingua-franca of global business. Journal of emerging technologies and innovative research, 6(2).
4. Anistasya, C. (2022). Students' difficulty in writing English: Affected by vocabulary skills. Journal of English Education, Literature and Linguistics, 5(2), 42–47. https://doi.org/10.31540/jeell.v5i2.1521
5. Brown, H. D. (2007). Principles of language learning and teaching. USA: Longman
6. Council of Europe (2018). Common European framework of reference for languages: Learning, teaching, assessment, companion volume with new descriptors. Strasbourg Cedex: Council of Europe.
7. Dahri et al. (2024) Investigating AI-based academic support acceptance and its impact on students' performance in Malaysian and Pakistani higher education institutions. Education and Information Technologies. https://doi.org/10.1007/s10639-024-12599-x
8. Dewi, U. (2023). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. Lingua Cultura, 16(2), 155–161. https://doi.org/10.21512/lc.v16i2.8315
9. Fahmi, M. & Cahyono, B. (2021). EFL students' perception on the use of Grammarly and teacher feedback. Jees (Journal of English Educators Society), 6(1), 18–25. https://doi.org/10.21070/jees.v6i1.849
10. Faisal, F., & Carabella, P. A. (2023). Utilizing Grammarly in an academic writing process: Higher-Education students' perceived views. Journal of English Language Teaching and Linguistics, 8(1), 23. https://doi.org/10.21462/jeltl.v8i1.1006
11. Fitri, A. I., Masrul, M., & Asilestari, P. (2022). An analysis on students' ability in writing descriptive text. Journal of English Language and Education, 7(2), 105–112. https://doi.org/10.31004/jele.v7i2.290
12. Hadiat, A. W. F. (2022). The use of Grammarly to enhance students' accuracy in writing descriptive text. Journal of English Education Program (JEEP), 9(2), 133. https://doi.org/10.25157/(jeep).v9i2.8552

13. Jerbi, D. (2023). Exploring the latest frontiers of artificial intelligence: A Review of trends and developments. TechRxiv. https://doi.org/10.36227/techrxiv.22717327.v1
14. Lestari, I. (2020). Error analysis of simple present in writing descriptive text. Eltics: Journal of English Language Teaching and English Linguistics, 5, 43-49. https://doi.org/10.31316/eltics.v5i2.748
15. Malaysian Examinations Council. (2019). Regulations, and test specifications. 1–45.
16. Manolescu, D. (2023). A quick snapshot of the English language. Journal of Critical Studies in Language and Literature, 4(1), 14–20. https://doi.org/10.46809/jcsll.v4i1.191
17. Maulidiyah, T. N., & Mandarani, V. (2023). Deciphering descriptive text challenges: Seventh grade students' writing difficulties. Academia Open, 8(1). https://doi.org/10.21070/acopen.8.2023.3020
18. Mubarok et al. (2020). Grammarly: An online EFL writing companion. Eltics, 5(2).
19. Naka, L., & Spahija, D. (2022). Impact of English language as a human capital in the higher education institutions' development strategy. Corporate and Business Strategy Review, 3(2), 262–272. https://doi.org/10.22495/cbsrv3i2siart7
20. Osamor et al. (2023). Chatbots and AI in education (AIEd) tools: The good, the bad, and the ugly. Journal of Applied Learning and Teaching, 6(2), pp. 332–345. https://doi.org/10.37074/jalt.2023.6.2.29
21. Pavlyuk, E. S., & Salisu, C. A. (2022, June 13). English as a tool for cross cultural interaction in business education. Chronos, 7(4), 96–99. https://doi.org/10.52013/2658-7556-66-4-28
22. Rahim et al. (2022). AI-based chatbots adoption model for higher-education institutions: A hybrid PLS-SEM-neural network modelling approach. Sustainability, 14(19), 12726. https://doi.org/10.3390/su141912726
23. Shahzod et al. (2021). English is a global language and importance of knowing foreign languages. Science Education, 2(6):578-580.
24. Sharma, P. (2023). The use of imagery and its significance in literary studies. Journal of English Studies, 14, 114–127. https://doi.org/10.3126/ojes.v14i1.56664
25. Singgih, M., & Rachmasisca, M. F. (2020). The effect of using the concept mapping model and student learning activities on narrative writing ability at SD Negeri 3 Selamider Bandar Lampung. Journal of Didactique Indonesian, 1(2), 3.
26. Singh, S.V., & Hiran, K.K. (2022). The impact of AI on teaching and learning in higher education technology. Journal of Higher Education Theory and Practice, 22(13), pp. 135–148. https://doi.org/10.33423/jhetp.v22i13.5514
27. Siregar, D. Y. (2022). Students' abilities in writing descriptive text at grade seventh students of SMP Dar Al-Falah in Tanjungbalai. Vision, 18(2), 130. https://doi.org/10.30829/vis.v18i2.2180
28. Solikhah, S. W., Turohmah, N. N., & Heriyanto, D. (2023). Student learning comparison using the song "Dear God" by Avenged Sevenfold to improve listening skills. Journal on Education, 6(1), 710–718. https://doi.org/10.31004/joe.v6i1.2985
29. Wulva, K. S. (2023). Students' ability to rewrite the retelling descriptive text. International Journal of Business and Information Technology, 4(1), 32–41. https://doi.org/10.47927/ijobit.v4i1.660
30. Younis et al. (2023). ChatGPT Evaluation: Can it replace Grammarly and Quillbot tools? British Journal of Applied Linguistics, 3(2), 34–46. https://doi.org/10.32996/bjal.2023.3.2.4