Evaluating the Use of Grammarly to Improve Academic Writing: A Dual Assessment among ICT Students

Ratna Widya Iswara[1*], Fitria Nugrahani[2], Dewi Kurniawati[3], Wafana Mumtaza Dzauqiyah[4], Ahmad Dandy Dwi Nugroho[5]

[1,2,3,4,5] Politeknik Negeri Jakarta, Depok 16425, Indonesia
ratna.widya.iswara@tik.pnj.ac.id

Abstract. Academic writing remains a persistent challenge for students in vocational ICT programs, particularly in balancing technical content with linguistic accuracy. This study evaluates the effectiveness of Grammarly as an AI-based writing support tool by employing a dual-assessment approach that combines automated scoring and human evaluation. Seventeen third-year ICT students enrolled in an English Proficiency Test Preparation course completed two IELTS-style Writing Task 2 essays in a quasi-experimental pre-test–post-test design. Each writing sample was assessed using Grammarly's automated score and the IELTS Writing Task 2 Band Descriptors applied by three independent English raters. The results indicate a modest improvement in Grammarly scores (mean gain: +3.8) and a slight increase in IELTS overall band scores (mean gain: +0.1), with the most notable improvement observed in Grammatical Range and Accuracy. However, discrepancies between automated and human scores were identified, and minor score variations occurred when identical texts were resubmitted to Grammarly, raising concerns regarding scoring stability. These findings suggest that while Grammarly is effective in enhancing surface-level linguistic accuracy, its contribution to overall writing quality remains limited. Consequently, Grammarly should be integrated as a supplementary pedagogical tool rather than a standalone assessment method. Future research may further explore learners' engagement with automated feedback and the pedagogical integration of AI-based writing tools in academic writing instruction.

Keywords: Academic Writing, Grammarly, ICT Students, IELTS.

1 Introduction

Writing skills play a vital role in both academic and professional contexts, serving as a strong foundation for effective communication, critical thinking, and knowledge dissemination [1]. In vocational higher education, particularly those in non-language fields such as Information and Communication Technology (ICT), mastering academic writing often presents a significant challenge [2, 3]. These students are required not only to master discipline-specific technical concepts but also to communicate them accurately and coherently in English, a task that is frequently hindered by limited linguistic proficiency and restricted exposure to formal academic writing instruction.

In response to these challenges, technology-based language support tools have increasingly been adopted to assist students in improving their writing. [4], for example,

M. A. Norazlina et al. (eds.), *Proceedings of the 4th International Conference of Research on Language 2025 (IROLE 2025)*, Advances in Social Science, Education and Humanities Research 1009,
https://doi.org/10.2991/978-2-38476-563-8_4

conducted a comparative study of several AI-powered writing assistance tools, and found that Grammarly has become one of the widely recognized tools for its grammar and spelling corrections, which focuses more on enhancing language clarity, while also providing reasonably effective plagiarism detection. In addition to large language models such as ChatGPT and DeepSeek, Grammarly has also evolved into an AI-powered writing assistant, combining traditional automated error detection with generative AI features to support language learning and academic writing [5]. Previous studies have demonstrated that such tools can help reduce surface-level language errors and increase learners' confidence in writing [6, 7].

However, existing research on AI-based writing tools, including Grammarly, has predominantly focused on general writing improvement, technical performance, or learner perceptions [8]. Empirical evidence measuring actual writing development through standardized and externally validated assessment frameworks remains limited. In particular, few studies have examined whether Grammarly's automated feedback contributes to measurable improvement in overall writing quality beyond grammatical accuracy, especially when evaluated using internationally recognized rubrics such as the IELTS Writing Band Descriptors.

This gap is especially evident in vocational ICT contexts, where students frequently rely on digital tools but often receive minimal structured training in academic writing. While Grammarly is widely used by students worldwide [9], its pedagogical effectiveness for ICT students who must balance technical content with academic language demands, remains underexplored. Moreover, little attention has been given to comparing automated writing feedback with human evaluation to capture both micro-level linguistic improvements and macro-level communicative competence.

To address these gaps, the present study evaluates the effectiveness of Grammarly as a writing support tool for ICT students. Students' writing performance is assessed using a dual-assessment approach, combining Grammarly's automated scoring with human evaluation based on the IELTS Writing Band Descriptors. By integrating automated and human perspectives, this study aims to provide a more comprehensive understanding of Grammarly's role in supporting academic writing development in vocational ICT education.

2 Literature Review

2.1 Automated Writing Evaluation (AWE) Tools in EFL Contexts

Automated Writing Evaluation (AWE) tools have increasingly been integrated into EFL and ESP classrooms to support learners' writing development, particularly in terms of grammatical accuracy, lexical choice, and text clarity. Previous studies have reported that AWE tools can promote learner autonomy by enabling immediate feedback and iterative revision processes [5, 9].

Despite these advantages, a growing body of research highlights inherent limitations in AWE-based feedback. While automated systems are effective in identifying surface-level linguistic errors, they tend to provide limited support for higher-order writing skills such as argument development, task fulfillment, and discourse coherence. As a

result, improvements facilitated by AWE tools are often confined to micro-level linguistic features rather than holistic writing quality.

Within this broader AWE landscape, Grammarly has emerged as one of the most widely adopted writing support tools among EFL learners. Studies report that students generally perceive Grammarly as a useful tool for revising grammatical errors, improving sentence-level accuracy, and enhancing textual clarity [10]. Empirical quasi-experimental research has further demonstrated measurable improvements in writing accuracy following Grammarly use, particularly in grammar-related aspects [11].

These findings are further supported by Wang, who reported that university-level English essays did not show significant improvement in lexical sophistication after Grammarly-based correction, as the tool primarily provides automated evaluations targeting general writing quality rather than in-depth lexical or discourse development [12]. Collectively, these studies suggest that while Grammarly is effective in supporting surface-level revisions, its pedagogical impact on comprehensive academic writing proficiency remains constrained.

2.2 Human-Based Writing Assessment and the Role of IELTS Rubrics

The limitations of AWE tools underscore the continued importance of human-based writing assessment frameworks that evaluate writing holistically. Standardized rubrics such as those used in the International English Language Testing System (IELTS) provide a more comprehensive evaluation of writing proficiency by incorporating both linguistic accuracy and communicative effectiveness. IELTS is widely recognized as a global standard for assessing English language proficiency and is valued for its reliability, validity, and consistency across diverse educational contexts.

In the IELTS Writing component, performance is assessed using four criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy, each contributing equally to the overall band score. Unlike automated scoring systems, IELTS assessment emphasizes argument relevance, idea development, discourse organization, and lexical appropriacy in relation to task demands. Consequently, IELTS-based evaluation captures dimensions of writing quality that are not fully addressed by AWE tools such as Grammarly.

Given these differences, comparing Grammarly's automated feedback with IELTS-based human assessment offers valuable insights into the extent to which automated tools can meaningfully support academic writing development. Such comparison enables researchers to distinguish between improvements in surface-level linguistic accuracy and broader gains in writing proficiency. This distinction is particularly important in EFL and ESP contexts, where students may demonstrate improved grammatical correctness without corresponding advancement in discourse-level writing skills.

Although existing studies have documented Grammarly's effectiveness in improving grammatical accuracy and learner confidence, empirical research employing dual-assessment frameworks that combine automated feedback with standardized human evaluation remains limited. Furthermore, few studies have focused on vocational ICT students, who face unique challenges in balancing technical content with academic language demands. To address these gaps, the present study examines Grammarly's effectiveness using both automated scoring and IELTS Writing Task 2

descriptors, providing a more comprehensive evaluation of its pedagogical role in academic writing instruction.

3 Method

This study employed a quasi-experimental, one-group pre-test–post-test design to investigate the effect of Grammarly on academic writing quality among ICT students. The participant were 17 third-year students enrolled in an English Proficiency Test Preparation course. According to an initial survey, these students had never taken an official IELTS test, but they were familiar with IELTS assessment criteria and strategies, which had been explicitly taught as part of the course.

Participants completed two IELTS-style Writing Task 2 essays as a pre-test and a post-test. The writing prompts addressed technology-related issues within an ICT context and were designed to be comparable in terms of task complexity, argumentative structure, and cognitive demand, thereby minimizing topic-related bias. The pre-test task discussed the positive and negative impacts of technology on society, while the post-test task explored a similar theme with a different framing to avoid memorization effects.

After completing the pre-test, students received brief instructional guidance on how to interpret Grammarly feedback for self-revision, focusing on grammar accuracy, coherence, clarity, and lexical choice. The post-test essay was then completed using Grammarly as a revision support tool.

Each writing task was evaluated using a dual-assessment approach: (1) automated feedback and scoring generated by Grammarly (web-based version), and (2) human evaluation based on the IELTS Writing Band Descriptors (Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy). The human assessment was conducted by three English lecturers with experience in IELTS instruction and assessment.

To ensure scoring consistency, inter-rater reliability among the three human raters was examined using Intraclass Correlation Coefficient, indicating an acceptable level of agreement. The final IELTS writing score for each essay was obtained by averaging the scores from the three raters.

The use of both automated and human scoring enabled the analysis of writing improvement from micro-level linguistic accuracy and macro-level communicative effectiveness, providing a more comprehensive evaluation of Grammarly's pedagogical impact.

4 Research Findings

This study aimed to evaluate the effectiveness of Grammarly in improving the academic writing quality of ICT students by comparing pre-test and post-test results using automated Grammarly scores and human evaluation based on IELTS Writing Task 2 descriptors. The descriptive statistics of students' writing performance are presented in Table 1.

Table 1. Pre-test and Post-test Writing Scores.

Indicator	Pre-Test Mean	Post-Test Mean	Mean Gain (Δ)
Grammarly Score	64.0	67.8	+3.8
IELTS Overall Band	6.5	6.6	+0.1
Task Response (TR)	6.6	6.7	+0.1
Coherence and Cohesion (CC)	6.5	6.6	+0.1
Lexical Resource (LR)	6.6	6.6	+0.1
Grammatical Range and Accuracy (GRA)	6.2	6.4	+0.2

As shown in Table 1, students' Grammarly scores increased from a mean of 64.0 in the pre-test to 67.8 in the post-test, resulting in a mean gain of +3.8 points. Based on human evaluation using the IELTS Writing Band Descriptors, the overall writing band score increased marginally from 6.5 to 6.6.

Among the four IELTS criteria, the highest gain was observed in Grammatical Range and Accuracy (GRA), with an increase of +0.2. The remaining criteria, which are Task Response, Coherence and Cohesion, and Lexical Resource, showed only minimal improvement (+0.1) or no change. Overall, the results indicate a greater improvement in grammar-related aspects compared to higher-order writing components.

In addition to mean score comparisons, an individual-level analysis was conducted to examine variations in students' IELTS writing performance deeply. The number and percentage are shown in Table 2.

Table 2. Distribution of Individual Changes in IELTS Overall Writing Scores

Change Category	Number of Students	Percentage
Improved (≥ +0.5 band)	6	35.3%
No Change (0.0)	4	23.5%
Decline (≤ −0.1)	7	41.2%
Total	**17**	**100%**

As presented in Table 2, six students (35.3%) demonstrated improvement of at least 0.5 band in their overall IELTS writing score. Four students (23.5%) showed no change, while seven students (41.2%) experienced a slight decline in their post-test scores. These results indicate that Grammarly's impact on writing performance varied considerably across individuals.

5 Discussion

The findings indicate that Grammarly contributed more noticeably to improvements at the micro-level of writing, particularly in grammatical accuracy, than to macro-level writing quality as reflected in the IELTS overall band score. The relatively larger gain in Grammarly scores compared to IELTS scores suggests that automated feedback is

more sensitive to surface-level linguistic changes, such as grammar, mechanics, and sentence clarity. This finding aligns with previous research suggesting that automated writing evaluation (AWE) tools can effectively support error detection and improve grammatical accuracy [5].

The improvement observed in Grammatical Range and Accuracy (GRA) aligns with Grammarly's primary function as an automated writing evaluation tool designed to identify grammatical errors and suggest syntactic revisions. This result supports previous findings that AWE tools are effective in enhancing lower-level language features but have limited impact on higher-order writing skills such as idea development, coherence, and lexical sophistication [12]. In contrast, minimal gains in Lexical Resource and Coherence and Cohesion indicate that Grammarly's feedback may not sufficiently support students in expanding vocabulary use or organizing arguments more effectively.

The modest improvement in Task Response further highlights a gap between Grammarly's capabilities and the demands of academic writing tasks such as IELTS Writing Task 2, which require clear stance-taking, idea elaboration, and relevance to the prompt. Since Grammarly does not provide explicit guidance on argument quality or content relevance, students may improve sentence-level accuracy without substantially improving their overall communicative effectiveness.

These findings underscore the importance of positioning Grammarly as a complementary pedagogical tool rather than a standalone solution. For instructors, Grammarly can be effectively integrated during the drafting and revision stages to raise students' awareness of grammatical accuracy and encourage self-editing practices. However, teacher-led instruction remains essential for developing higher-order writing skills, including argument structure, coherence, and task interpretation. Combining Grammarly feedback with targeted classroom instruction, peer review activities, and rubric-based discussions may help students better connect linguistic accuracy with overall writing quality.

While the tool can enhance grammatical awareness and revision practices, writing development in an academic context still requires teacher guidance, peer feedback, and explicit instruction in discourse organization. This aligns with existing critiques of AWE systems, which argue that these tools are best used as complementary aids rather than replacements for human feedback [8].

The individual variation observed in Table 2 suggests that Grammarly's effectiveness is not uniform across learners. While a substantial proportion of students demonstrated measurable improvement, others showed stable or declining performance. This pattern may be attributed to differences in initial proficiency levels, writing strategies, and students' ability to interpret and apply automated feedback meaningfully. In some cases, an increased focus on grammatical correction may have diverted attention from content development and task fulfillment, which are critical components of IELTS Writing Task 2.

These findings suggest that Grammarly should be integrated selectively rather than uniformly. Teachers may consider guiding lower-proficiency students more explicitly in interpreting feedback, while higher-proficiency students may benefit from using Grammarly as a self-editing tool. Structured scaffolding and post-feedback reflection activities are therefore essential to maximize Grammarly's pedagogical value.

An additional observation emerged during data processing: when identical texts were resubmitted to Grammarly at different times, slight score variations (within 1–2 points) were occasionally observed. Although these differences did not alter the overall trend, they suggest that Grammarly's scoring algorithm may not produce entirely stable results for identical inputs. This variability supports existing concerns regarding the reliability of AI-based scoring systems, particularly in research or high-stakes assessment contexts [5]. Consequently, Grammarly's scores should be interpreted cautiously and supplemented with human evaluation when used for formal assessment purposes.

6 Conclusion

This study investigated the effectiveness of Grammarly as a writing support tool for ICT students by employing a quasi-experimental pre-test–post-test design and a dual-assessment framework combining automated Grammarly scores and human evaluation based on IELTS Writing Task 2 descriptors. The findings indicate that Grammarly contributed to measurable improvement in grammar-related aspects of students' writing, as reflected in both automated scores and the Grammatical Range and Accuracy criterion of the IELTS assessment.

However, improvements in overall writing quality, particularly in higher-order components such as task response, coherence, and lexical resource, remained modest. The discrepancy between the relatively larger gains identified by Grammarly and the limited changes observed through human evaluation suggests that automated feedback tools are more sensitive to surface-level linguistic revisions than to discourse-level writing development. Additionally, minor inconsistencies in Grammarly scoring across repeated submissions highlight limitations in relying on automated systems as standalone assessment instruments.

From a pedagogical perspective, these findings suggest that Grammarly should be positioned as a supplementary tool to support self-editing and grammatical awareness rather than as a replacement for teacher feedback or standardized writing assessment. Effective integration of Grammarly in academic writing instruction requires structured guidance, reflective activities, and complementary human feedback to address higher-order writing skills.

This study is limited by its small sample size and short intervention duration, which may restrict the generalizability of the findings. Future research could examine learners' cognitive engagement with automated feedback, investigate differential effects across proficiency levels, or compare the pedagogical effectiveness of multiple AI-based writing tools using more controlled experimental designs.

Acknowledgments. This research is fully funded by Politeknik Negeri Jakarta through Agreement Letter 853/PL3.A.10/PT.00.06/2025, dated June 2, 2025. The authors send a sincere appreciation and thanks to all parties who have contributed to this research.

Disclosure of Interests. The authors have research grant from Politeknik Negeri Jakarta (Agreement Letter 853/PL3.A.10/PT.00.06/2025)

References

1. Taye, T., Mengesha, M.: Identifying And Analyzing Common English Writing Challenges Among Regular Undergraduate Students, Heliyon, **10**(17), Sep. (2024)
2. Yang, W., Wen Y.: International Journal of Education and Humanities Writing Strategies and Writing Proficiency of Chinese Vocational College Students
3. Konstantinidou, L., Madlener-Charpentier, K., Opacic, A., Gautschi, C., Hoefele, J.: Literacy In Vocational Education And Training: Scenario-Based Reading And Writing Education, Read Writ, **36**(4), pp. 1025–1052 (2023)
4. Tamilselvi, C., Dhanasakkaravarthi, B., Devi, R., Ebenezer, A.S., Dhanwanth, B., Mohanaprakash, T.A.: A Comprehensive Study of AI-Powered Writing Assistance Tools and Content Spinning Applications, in 7th International Conference on Electronics, Communication and Aerospace Technology, ICECA Proceedings, Institute of Electrical and Electronics Engineers Inc., pp. 1104–1111 (2023)
5. Ranalli, J.: Automated Written Corrective Feedback: How Well Can Students Make Use Of It?, Comput Assist Lang Learn, **31**(7), pp. 653–674 (2018)
6. Zunaidah, A., Wiharja, C. K., Wicaksono, D.W.: User Experience with Grammarly's Generative AI: Ethical Implications for Improving Writing Skills, in Proceedings - International Conference on Education and Technology, ICET, Institute of Electrical and Electronics Engineers, pp. 86–91 (2024)
7. Damayanti, S., Azizah, S.: Grammarly's Effectiveness in Enhancing English Writing: A Case Study of Vocational High School Students, PANYONARA: Journal of English Education, **6**(2), pp. 15–24 (2024)
8. Huang, H.W., Li, Z., and Taylor, L.: The Effectiveness of Using Grammarly to Improve Students' Writing Skills, in ACM International Conference Proceeding Series, Association for Computing Machinery, pp. 122–127 (2020)
9. Karatay, Y., Karatay, L.: Automated Writing Evaluation Use In Second Language Classrooms: A Research Synthesis, Jul. 01, Elsevier Ltd (2024)
10. Miranty, D., Widiati, U., Cahyono, B. Y. and Sharif, T. I. S. T.: Automated Writing Evaluation Tools For Indonesian Undergraduate English as a Foreign Language Students' Writing, International Journal of Evaluation and Research in Education, **12**(3), pp. 1705–1715 (2023)
11. Hidayat, A., Raden, U., Lampung, I., Sari, R.: The Effect of Using Grammarly on EFL Students' Writing Accuracy: A Quasi-Experimental Research [Online]. Available: http://ejournal.radenintan.ac.id/index.php/JLSS/index://creativecommons.org/licenses/by-sa/4.0 (2025)
12. Wang, C.: Analyzing Composition Complexity in Revised Versions through Artificial Intelligence-Based Writing Assistants, in Proceedings 9th Annual International Conference on Network and Information Systems for Computers, ICNISC 2023, Institute of Electrical and Electronics Engineers Inc., pp. 104–106 (2023)