Automated writing feedback with grammarly: An experimental study among medical students at a university in Vietnam

Do Phuong Linh Chi[1]*, Pham Bich Ngoc[1], Nguyen Thu Hien[1]

ABSTRACT

Academic writing in English is essential for medical students as it enables access to global resources and scholarly publication (Hyland & Hyland, 2019). Yet many Vietnamese learners encounter persistent difficulties with grammar, vocabulary, and textual coherence, which restrict their academic performance. In response, Automated Writing Corrective Feedback (AWCF) systems such as Grammarly have emerged, offering immediate, technology-supported assistance. While widely adopted, their pedagogical contribution in medical education remains underexplored (Ghufron, 2019; Dizon & Gayed, 2021). This experimental study investigates Grammarly's role in supporting academic writing among medical students at Hai Phong University of Medicine and Pharmacy. Sixty second-year students were randomly allocated to an experimental group using Grammarly Premium or a control group receiving teacher feedback. Writing assignments before and after the intervention were assessed by certified IELTS examiners, and semi-structured interviews with five instructors were thematically analyzed. Findings showed that both groups improved, Statistical analysis using paired-samples t-tests revealed significant improvements within both groups., but teacher feedback produced significantly greater gains. Teachers valued Grammarly for efficiency and learner autonomy, yet noted risks of over-reliance and limited capacity to address higher-order writing issues. The study concludes that Grammarly should supplement, not replace, teacher feedback in medical English writing instruction.

Keywords: Grammarly, automated feedback, academic writing, medical students, Vietnam

[1] *Hai Phong University of Medicine and Pharmacy, Vietnam*

* Corresponding author

Do Phuong Linh Chi
Email: dplchi@hpmu.edu.vn

INTRODUCTION

Background and Rationale for the Research

Mastery of English academic writing is increasingly indispensable for medical students, as it provides access to international literature, participation in scholarly dialogue, and opportunities for publication (Hyland & Hyland, 2019). Despite this importance, Vietnamese learners of English frequently struggle with grammar, vocabulary choice, and the development of coherent arguments, leading to limitations in producing discipline-specific texts (Nguyen, 2020).

Corrective feedback has long been recognized as a cornerstone of second language writing instruction. Teacher feedback, in particular, has been shown to promote accuracy, improve text organization, and foster critical thinking skills (Ferris, 2010; Bai & Hu, 2017). Nevertheless, the growing adoption of technology in education has introduced new tools that complement traditional

approaches. Among these, Automated Writing Corrective Feedback (AWCF) tools such as Grammarly have gained prominence, offering instant error detection and revision suggestions for grammar, vocabulary, and style (Ghufron, 2019).

While several studies highlight the benefits of AWCF in improving surface-level accuracy and supporting learner autonomy (Li, Link, & Hegelheimer, 2015; Dizon & Gayed, 2021), concerns remain regarding their limitations. Research indicates that these tools often fail to provide effective feedback on higher-order writing features such as argumentation, coherence, and discipline-specific terminology (O'Neill & Russell, 2019; Dodigovic & Tovmasyan, 2021). Such shortcomings are particularly relevant in medical education, where precision and clarity are crucial.

Given these gaps in the existing literature, the research addresses two main questions:

(1) Does Grammarly improve medical students' academic writing performance compared to teacher feedback?

and (2) How do instructors perceive the pedagogical value of Grammarly?

To address these questions, this study investigates the effectiveness of Grammarly in supporting medical students' academic writing at Hai Phong University of Medicine and Pharmacy by comparing writing outcomes between students using Grammarly and those receiving teacher feedback, while also examining instructors' perceptions of the tool's pedagogical value.

The Role of Feedback in Academic Writing

Feedback is widely acknowledged as an essential component in academic writing instruction, helping learners notice their errors, improve accuracy, and develop rhetorical competence. Hyland and Hyland (2019) emphasize that timely and constructive feedback not only corrects surface-level errors but also fosters students' critical thinking and revising strategies. In medical education, where clarity and accuracy are indispensable for producing case reports, research articles, and scientific abstracts, the importance of feedback becomes even more evident (Nguyen, 2020). However, teacher-provided feedback is often constrained by class size and time limitations, leading to delays and inconsistency (Mao & Crosthwaite, 2019). This has prompted the exploration of technology-driven alternatives.

Automated Writing Corrective Feedback (AWCF)

Automated Writing Corrective Feedback (AWCF) refers to computer-mediated tools that automatically detect writing errors and provide corrective suggestions to language learners. Such systems are widely valued for their immediacy, accessibility, and scalability, as they allow students to receive feedback instantly without the time constraints typically associated with teacher-mediated correction (Bai & Hu, 2017). By offering continuous and on-demand support, AWCF tools can facilitate autonomous learning and encourage learners to take greater responsibility for revising their own writing. Empirical studies have shown that AWCF can reduce teachers' workload by handling repetitive surface-level errors, thereby enabling instructors to allocate more time to higher-order pedagogical concerns (Li, Link, & Hegelheimer, 2015). Moreover, the availability of automated feedback supports multiple rounds of self-revision, which is a key feature of process-oriented approaches to writing instruction.

Despite these advantages, AWCF tools have been subject to sustained criticism in the

literature. A central concern is that such systems tend to focus predominantly on grammatical accuracy and lexical choice, while offering limited support for higher-order aspects of writing, including cohesion, coherence, argument development, and rhetorical organization (Ferris, 2010). As a result, improvements facilitated by AWCF may remain largely at the micro-linguistic level and may not translate into broader gains in overall writing quality. In addition, O'Neill and Russell (2019) caution that over-reliance on automated feedback may encourage students to accept corrections uncritically, particularly when the system's suggestions are perceived as authoritative. This lack of critical engagement with feedback may constrain the development of deeper writing skills and metalinguistic awareness, highlighting the need for pedagogically guided use of AWCF in academic writing contexts.

Grammarly as a Leading AWCF Tool

Grammarly is one of the most widely used AWCF platforms in higher education and has gained considerable popularity among second language learners due to its user-friendly interface and immediate feedback. The tool provides automated feedback on a range of language features, including grammar, punctuation, spelling, vocabulary, and stylistic choices, making it particularly attractive for students engaged in academic writing. Empirical evidence suggests that Grammarly can contribute to improvements in learners' linguistic accuracy. For instance, Ghufron (2019) reported a significant reduction in grammatical errors among EFL students who incorporated Grammarly into their revision process. Similarly, Dizon and Gayed (2021) found that Grammarly enhanced the overall quality of mobile-assisted L2 writing tasks, particularly in terms of language form and clarity.

Despite these reported benefits, the literature also points to important limitations in Grammarly's pedagogical effectiveness. Dodigovic and Tovmasyan (2021) observed that Grammarly frequently misclassified or overlooked errors that require contextual interpretation or discipline-specific knowledge, thereby limiting its usefulness in specialized academic domains. O'Neill and Russell (2019) further noted that writing advisors expressed concern that Grammarly may encourage "surface-level editing" by directing learners' attention primarily to form rather than meaning or rhetorical purpose. More critically, Iskender (2023) warned that students' over-reliance on Grammarly could undermine the development of independent editing skills and critical thinking, as learners may defer evaluative judgment to the automated system rather than engaging reflectively with their own writing. These findings suggest that while Grammarly is a powerful AWCF tool for supporting language accuracy, its pedagogical value depends largely on how it is integrated into instructional practices and the extent to which learners are guided to use its feedback critically.

Grammarly in Discipline - Specific Contexts

Most studies on Grammarly focus on general ESL/EFL contexts. However, discipline-specific writing presents additional challenges. In medical education, students must use precise terminology, adhere to research conventions, and produce texts that demonstrate both linguistic accuracy and subject-matter expertise. Nguyen (2020) highlighted that Vietnamese medical students often face persistent problems with

grammar and vocabulary when writing research abstracts, which negatively affect their ability to disseminate findings internationally. While Grammarly can address low-level linguistic errors, it is less effective in supporting higher-order skills essential for medical academic writing, such as structuring arguments and synthesizing evidence (Ferris, 2010).

Despite Grammarly's growing popularity, there is limited empirical evidence on its pedagogical value in medical English contexts. To date, no large-scale experimental studies in Vietnam have systematically compared Grammarly's effectiveness with traditional teacher feedback among medical students. This represents a critical research gap.

The existing literature highlights three key strands relevant to the present study. First, feedback is widely recognized as indispensable to the development of academic writing, with automated written corrective feedback (AWCF) increasingly valued for its efficiency, immediacy, and accessibility in supporting learners' revision processes (Hyland & Hyland, 2019; Bai & Hu, 2017). Second, empirical studies have demonstrated that Grammarly can enhance linguistic accuracy in general EFL writing contexts (Ghufron, 2019; Dizon & Gayed, 2021); however, research also points to notable limitations, particularly in terms of contextual appropriateness, disciplinary sensitivity, and the potential risk of over-reliance on automated feedback (O'Neill & Russell, 2019; Dodigovic & Tovmasyan, 2021). Third, despite the growing body of research on AWCF tools, there remains a paucity of empirical studies examining the use of Grammarly in discipline-specific contexts, especially within medical English education in Vietnam, thereby underscoring

the need for contextually grounded investigations such as the present study (Nguyen, 2020).

Therefore, the present study investigates the effectiveness of Grammarly as an AWCF tool among medical students at Hai Phong University of Medicine and Pharmacy. By comparing Grammarly-based feedback with teacher feedback and analyzing instructors' perceptions, this research seeks to contribute evidence to guide the integration of AWCF into medical English writing instruction.

MATERIALS AND METHODS

Research Design

This study employed a quasi-experimental mixed-methods design to investigate the effectiveness of Grammarly as an Automated Writing Corrective Feedback (AWCF) tool in supporting academic writing among medical students. A mixed-methods approach was chosen to capture both the quantitative outcomes of students' writing performance and the qualitative insights from instructors' perceptions, thereby enabling a more comprehensive understanding of the pedagogical impact of AWCF (Creswell & Plano Clark, 2018).

The experiment was conducted over 12 weeks during the 2024–2025 academic year at Hai Phong University of Medicine and Pharmacy. Two intact classes enrolled in the English for Academic Purposes (EAP) course were assigned as the experimental group (EG) and the control group (CG) to maintain instructional authenticity. While both groups completed similar writing tasks under the same instructional conditions, the EG students revised their work using Grammarly Premium, whereas the CG students received direct feedback from their instructors throughout the intervention period.

Research Subjects

The participants consisted of 60 second-year medical students aged 19 to 21. They were selected based on their enrollment in the English for Academic Purposes course, successful completion of prerequisite English modules, and voluntary consent to participate in the study. All participants demonstrated comparable levels of English proficiency, as determined by institutional placement tests aligned with the B1 (intermediate) level of the Common European Framework of Reference for Languages (CEFR). To ensure group equivalence, students in the experimental and control groups were drawn from intact classes with similar instructional backgrounds and exhibited no statistically significant differences in baseline writing proficiency prior to the intervention. They were randomly assigned into two groups of 30 students each:

- Experimental Group (EG): Students revised their essays with the aid of Grammarly Premium.
- Control Group (CG): Students submitted their essays and received teacher-provided corrective feedback.

Additionally, five instructors who were responsible for teaching academic writing courses were recruited for the qualitative phase of the study. They were experienced English lecturers with at least five years of teaching at medical universities.

Research Instruments

To ensure reliability and validity, the study used a combination of instruments:

Writing Tests (Pre-test and Post-test)

Both the experimental and control groups completed a 250-word academic essay under timed conditions (60 minutes) at the beginning of the course (Week 1) and at the end of the instructional period (Week 12).

The essay prompts were aligned with medical-related themes, such as public health issues or healthcare innovations, in order to ensure disciplinary relevance and to elicit domain-appropriate academic language use. Student writing performance was assessed using the IELTS Writing Task 2 analytic scoring descriptors, which evaluate four key dimensions: task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy. To enhance scoring objectivity and reliability, all scripts were independently rated by two certified IELTS examiners. Inter-rater reliability was subsequently calculated following established procedures in educational assessment research (Brown, 2015).

Grammarly Premium

For the experimental group (EG), Grammarly Premium served as the primary automated written corrective feedback tool throughout the intervention. Students were instructed to submit two versions of each writing assignment: an initial draft and a revised draft produced after engaging with Grammarly's feedback. This two-draft submission procedure enabled a systematic examination of the extent and nature of revisions attributable to Grammarly, thereby allowing the analysis of how the tool influenced students' revision behaviors and writing development over the course of the study.

Semi-structured Interview Guide

To explore instructors' perceptions of Grammarly within the instructional context, a semi-structured interview protocol was developed to guide data collection. The interview questions focused on instructors' perceived strengths and limitations of Grammarly as an automated written corrective feedback tool, its potential for

pedagogical integration into medical English writing courses, and its perceived impact on student learning and writing development. This flexible interview format allowed participants to elaborate on their experiences and provide in-depth reflections, while ensuring consistency across interviews for systematic qualitative analysis.

Data Collection

Data collection was carried out in two phases:

Quantitative Phase

In the quantitative phase of the study, all participants completed a pre-test essay in Week 1 in order to establish baseline levels of writing proficiency prior to the intervention. From Weeks 2 to 11, students participated in two major writing tasks designed to support the development of academic writing skills within the instructional programme. During this intervention period, students in the experimental group (EG) revised their drafts using Grammarly as an automated written corrective feedback tool, whereas students in the control group (CG) revised their work based on teacher-provided written feedback. This differential feedback approach allowed for a systematic comparison of the effects of automated versus human feedback on writing development. In Week 12, all participants completed a post-test essay to assess changes in writing performance following the intervention. Scores from the pre- and post-test essays were subsequently subjected to paired-samples t-tests to examine within-group improvements over time, as well as to compare the magnitude of progress between the experimental and control groups.

Qualitative Phase

In the qualitative phase of the study, semi-structured interviews were conducted with five instructors at the conclusion of the experimental period to explore their pedagogical experiences and perspectives regarding the use of Grammarly. Each interview lasted approximately 30–40 minutes and was audio-recorded with the participants' informed consent. The interview recordings were subsequently transcribed verbatim, and the resulting transcripts were systematically coded and analyzed using thematic analysis following the procedures outlined by Braun and Clarke (2012). This analytic approach enabled the identification of recurring patterns and themes related to instructors' perceptions of Grammarly's pedagogical role, its affordances and constraints, and its implications for feedback practices in academic writing instruction.

Ethical Considerations

The study was conducted in accordance with the ethical standards and research guidelines of Hai Phong University of Medicine and Pharmacy. Prior to data collection, all participants were fully informed of the purpose, procedures, and voluntary nature of the study, and written informed consent was obtained. Participants were assured that their identities would remain confidential and that all data would be anonymized and used solely for research purposes. To prevent any potential power imbalance or coercion, students were explicitly informed that their decision to participate or withdraw from the study at any stage would not affect their course grades or academic standing. All research procedures were designed to minimize potential risks and to ensure the protection of participants' rights and well-being throughout the study.

RESULTS

Quantitative Results

To examine the impact of Grammarly compared to teacher feedback, both the experimental group (EG) and control group (CG) completed pre-test and post-test essays. Table 1 shows that the two groups were nearly identical in their pre-test performance, indicating comparable proficiency levels at the outset.

Table 1. Pre-test Scores of Experimental Group (EG) and Control Group (CG)

Group	N	Mean	SD	Min	Max
EG	30	5.02	0.41	4.5	5.5
CG	30	5.08	0.38	4.6	5.6

As shown in Table 1, the experimental group (EG) and the control group (CG) demonstrated highly comparable levels of academic writing proficiency at the beginning of the study. The mean pre-test scores of the EG (M = 5.02, SD = 0.41) and the CG (M = 5.08, SD = 0.38) were closely aligned, with overlapping score ranges, indicating a relatively homogeneous sample in terms of baseline writing ability. The independent-samples t-test confirmed that there was no statistically significant difference between the two groups at the outset (t = 0.52, p > .05), suggesting that any subsequent differences observed in post-test performance could be more confidently attributed to the instructional interventions rather than pre-existing proficiency gaps.

Following the 12-week intervention period, both groups exhibited improvement in academic writing performance, indicating that sustained writing practice combined with feedback-whether automated or teacher-provided-contributed positively to students' writing development. This overall improvement across groups is consistent with previous research highlighting the central role of feedback in supporting second language academic writing. However, despite gains observed in both conditions, the control group, which received teacher feedback, outperformed the experimental group, which relied on Grammarly for revision support. This finding suggests that while Grammarly may facilitate certain aspects of writing development, particularly at the surface linguistic level, teacher feedback appears to offer a comparatively stronger impact on overall writing quality as assessed through the IELTS Writing Task 2 descriptors.

The superior performance of the control group may be attributed to the nature of teacher feedback, which is typically more responsive to discourse-level concerns, such as task achievement, coherence and cohesion, and the alignment of arguments with disciplinary expectations. In contrast, Grammarly primarily targets language accuracy and stylistic issues, which, although essential, may be insufficient on their own to substantially enhance higher-order writing skills within a relatively short instructional timeframe. This pattern of results indicates that automated written corrective feedback, while beneficial, may function more effectively as a complementary rather than a standalone feedback mechanism in academic writing instruction for medical students.

Importantly, the comparable baseline performance of the two groups strengthens the internal validity of the findings, as the observed post-intervention differences are unlikely to be confounded by initial proficiency disparities. Instead, the results point to differential pedagogical effects associated with the type of feedback provided, underscoring the continued

pedagogical value of teacher-mediated feedback in discipline-specific academic writing contexts.

Table 2. Post-test Scores of Experimental Group (EG) and Control Group (CG)

Group	N	Mean	SD	Min	Max
EG	30	5.63	0.39	5.1	6.2
CG	30	6.02	0.36	5.4	6.6

Table 2 presents the post-test writing performance of the experimental group (EG) and the control group (CG) following the 12-week intervention. Both groups demonstrated higher mean scores compared to their pre-test results, indicating overall improvement in academic writing performance over the course of the study. However, a clear difference emerged between the two groups at the post-test stage. The control group achieved a higher mean score (M = 6.02, SD = 0.36) than the experimental group (M = 5.63, SD = 0.39), with score distributions showing limited overlap. This result indicates that students who received teacher feedback outperformed those who relied primarily on Grammarly for revision.

The higher post-test performance of the control group suggests that teacher-provided feedback was more effective in fostering overall improvements in academic writing quality, as measured by the IELTS Writing Task 2 criteria. While Grammarly-supported revision contributed to measurable gains in the experimental group, the magnitude of improvement appears to be comparatively smaller. This pattern implies that automated written corrective feedback may be particularly effective in addressing surface-level language issues, whereas teacher feedback may offer stronger support for higher-order writing dimensions, such as task fulfillment, coherence and cohesion, and the development of well-structured arguments.

To further examine writing development over time, paired-samples t-tests were conducted to compare pre-test and post-test scores within each group. The results revealed statistically significant gains for both the experimental and control groups, confirming that sustained writing practice combined with feedback led to improved performance regardless of feedback type. Notably, however, the control group exhibited a larger mean gain from pre-test to post-test than the experimental group, reinforcing the finding that teacher feedback produced a stronger overall effect on writing development within the duration of the intervention.

Taken together, these results indicate that although Grammarly can function as a supportive tool for improving academic writing, particularly in facilitating revision and reducing linguistic errors, it does not fully replicate the pedagogical impact of teacher feedback in discipline-specific writing contexts. The superior performance of the control group highlights the importance of human-mediated feedback in guiding students' understanding of genre conventions, rhetorical organization, and disciplinary expectations in medical academic writing.

Table 3. Paired-Samples t-test Results (Pre-test vs Post-test, including p-values)

| Group	Mean Gain	t-value	p-value
EG	+0.61	4.87	< .05
CG	+0.94	6.12	< .01

Table 3 reports the results of the paired-samples t-tests comparing pre-test and post-test writing performance within each group. The findings indicate that both the experimental group (EG) and the control group (CG) demonstrated statistically significant improvements over the

12-week instructional period. Specifically, the EG achieved a mean gain of +0.61, with a t-value of 4.87 (p < .05), suggesting that the use of Grammarly contributed to meaningful improvements in students' academic writing performance. This result confirms that engagement with automated written corrective feedback can support writing development, particularly by facilitating revision and raising learners' awareness of linguistic accuracy.

However, the magnitude of improvement observed in the control group was notably larger. The CG recorded a mean gain of +0.94, accompanied by a higher t-value (t = 6.12, p < .01), indicating a stronger and more robust effect of teacher-provided feedback on writing development. The larger mean gain suggests that teacher feedback not only supported surface-level improvements but also more effectively addressed higher-order aspects of writing that are central to overall performance on the IELTS Writing Task 2 scale.

The difference in mean gains between the two groups (+0.94 for CG versus +0.61 for EG) is particularly important for interpreting the pedagogical effectiveness of Grammarly relative to teacher feedback. While both feedback conditions led to statistically significant progress, the greater gains achieved by the control group indicate that teacher feedback exerted a stronger impact on students' overall writing quality within the timeframe of the study. This finding implies that although Grammarly is beneficial as a feedback tool, its influence may be more limited when used as the primary source of feedback, especially in a discipline-specific academic writing context such as medical English.

From a pedagogical perspective, these results suggest that Grammarly may be most effective when positioned as a supplementary or pre-feedback tool that helps students address recurrent language errors prior to receiving teacher feedback. In contrast, teacher feedback appears better suited to guiding students' development of argumentation, organization, and task fulfillment-dimensions that typically require contextualized explanations and disciplinary expertise. The statistically significant gains observed in both groups nonetheless highlight the value of feedback-driven revision processes in academic writing instruction, while underscoring the continued central role of teachers in fostering higher-level writing development.
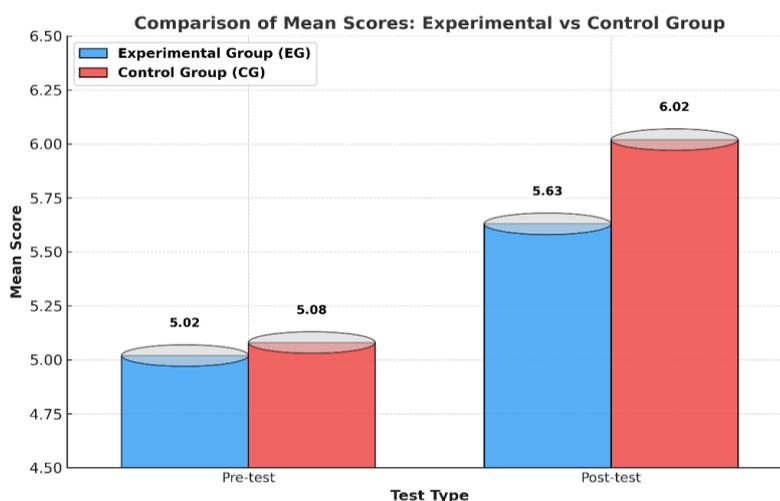
Figure 1. *Comparison of Mean Scores (EG vs CG)*

The chart visually reinforces the quantitative findings by illustrating the comparative mean scores of the experimental group (EG) and the control group (CG) at both the pre-test and

post-test stages. At the outset, the two groups exhibited nearly identical mean scores (EG = 5.02; CG = 5.08), confirming a comparable baseline level of academic writing proficiency. This close alignment at the pre-test stage visually supports the statistical evidence that no significant difference existed between the groups prior to the intervention, thereby strengthening the internal validity of the study.

In contrast, a clearer divergence emerges at the post-test stage. Although both groups demonstrated noticeable improvement, the control group achieved a higher mean score (6.02) than the experimental group (5.63). The visual gap between the two post-test bars highlights not only the presence of progress in both conditions but also the greater magnitude of improvement associated with teacher-provided feedback. This difference mirrors the paired-samples t-test results, which showed that the CG made larger gains than the EG over the 12-week period.

Importantly, the figure allows for an intuitive comparison of learning trajectories rather than isolated outcomes. While Grammarly-supported revision facilitated measurable progress for students in the experimental group, the steeper increase observed for the control group suggests that teacher feedback exerted a stronger overall influence on writing development. This pattern indicates that automated feedback may effectively support incremental improvements, particularly in language accuracy, but may be less impactful than teacher feedback in promoting broader gains in writing quality as captured by the IELTS Writing Task 2 scale.

Taken together, Figure 1 provides visual confirmation that feedback modality plays a critical role in shaping writing outcomes. The chart underscores the complementary value of Grammarly while simultaneously highlighting the continued pedagogical advantage of teacher-mediated feedback, especially in discipline-specific academic writing contexts such as medical English.

Qualitative Results

Interviews The semi-structured interviews with five instructors yielded rich insights into the pedagogical role of Grammarly in medical English writing instruction. Thematic analysis revealed three interrelated themes that illuminate both the affordances and limitations of Grammarly as an automated written corrective feedback tool within a discipline-specific context.

The first theme, efficiency and accessibility, highlights instructors' appreciation of Grammarly's capacity to provide immediate feedback on grammar, spelling, and basic language mechanics. Teachers consistently noted that this automated support reduced the need for repetitive, low-level corrections, thereby alleviating their feedback workload. As a result, instructors were able to redirect instructional time and cognitive resources toward higher-order writing concerns, such as content relevance, argument clarity, and organizational structure. As one instructor explained, "*Grammarly quickly deals with surface mistakes, so I can focus on content and structure.*" This finding suggests that Grammarly can function effectively as a pre-feedback or filtering tool, supporting more pedagogically meaningful teacher–student interactions.

The second theme, risks of over-reliance, reflects instructors' concerns regarding students' uncritical acceptance of Grammarly's suggestions. Several participants reported instances in which students adopted automated revisions without evaluating their semantic accuracy or disciplinary appropriateness. This issue was particularly salient in medical writing, where terminological precision and contextual meaning are crucial. As one instructor cautioned, "*Sometimes it changes a technical term, and students don't notice. That's risky in medical writing.*" These observations point to a potential pedagogical risk associated with AWCF use, namely the erosion of learners' critical engagement with language and content, especially when automated feedback is perceived as authoritative.

The third theme, limited support for higher-order writing, underscores the instructors' shared view that Grammarly provides little assistance with macro-level writing skills. All participants emphasized that the tool lacks the capacity to address key dimensions of academic writing development, including argumentation, coherence and cohesion, rhetorical organization, and the appropriate use of evidence within disciplinary conventions. As one instructor noted, "*It doesn't help students organize their ideas logically or use evidence properly.*" This limitation reinforces the distinction between surface-level linguistic accuracy and deeper rhetorical competence, suggesting that automated tools cannot substitute for teacher-mediated guidance in developing complex academic writing skills.

Taken together, these qualitative findings align closely with concerns raised in previous research, which has consistently shown that AWCF tools are effective in improving micro-level linguistic accuracy but remain insufficient for supporting macro-level writing development (Ferris, 2010; O'Neill & Russell, 2019). Importantly, the instructors' perspectives help explain the quantitative results of the present study, particularly the greater gains observed in the teacher-feedback condition. While Grammarly offers efficiency and immediacy, its pedagogical value appears to be constrained by students' levels of critical engagement and by the tool's limited capacity to address higher-order, discipline-specific writing demands.

DISCUSSION

The findings of the present study confirm the dual role of Grammarly in academic writing instruction and, more broadly, illuminate the theoretical and pedagogical limitations of automated written corrective feedback (AWCF) systems. While Grammarly demonstrates clear benefits as a supplementary feedback tool, the results also reveal its inadequacy as a stand-alone mechanism for supporting comprehensive academic writing development. These outcomes reinforce longstanding concerns that AWCF tools, by design, prioritize form-focused correction and lack the capacity to process contextual nuances or evaluate content quality in discipline-specific writing. On the one hand, the statistically significant gains achieved by the experimental group indicate that Grammarly can play a meaningful role in improving grammatical accuracy and lexical appropriacy. This finding is consistent with previous studies conducted in general EFL contexts, which report positive effects of Grammarly on surface-level language accuracy (Ghufron, 2019; Dizon & Gayed, 2021). Beyond linguistic improvement, Grammarly appeared to promote productive revision practices by encouraging multiple drafting and fostering a degree of learner autonomy. The immediacy and accessibility of

automated feedback reduced students' reliance on teachers for low-level error correction, potentially lowering anxiety and increasing opportunities for self-directed learning. From a process-oriented writing perspective, Grammarly thus functions as a facilitative tool that supports iterative revision and language awareness.

On the other hand, the superior performance of the control group underscores the irreplaceable pedagogical value of teacher-provided feedback, particularly in the development of higher-order writing skills. Teachers' feedback addressed dimensions such as coherence, argument development, rhetorical organization, and disciplinary accuracy-areas that remain beyond Grammarly's operational scope (Dodigovic & Tovmasyan, 2021; Iskender, 2023). In medical English writing, accuracy extends beyond grammatical correctness to include precise use of terminology, appropriate interpretation of evidence, and logical integration of ideas within disciplinary conventions. These complex meaning-making processes require contextual judgment and epistemic awareness, which automated systems are currently unable to provide. The greater gains observed in the teacher-feedback condition therefore suggest that human feedback is more effective in supporting the macro-level dimensions of academic writing assessed by the IELTS Writing Task 2 framework.

The qualitative findings further elucidate this quantitative pattern. Instructors consistently recognized Grammarly as a time-saving tool that efficiently handles repetitive surface-level errors, thereby allowing teachers to allocate more instructional time to higher-order concerns. At the same time, they expressed concern about students' tendency to uncritically accept automated suggestions, even when such revisions altered meaning or compromised disciplinary accuracy. These reflections lend empirical support to O'Neill and Russell's (2019) argument that Grammarly may inadvertently encourage surface-level editing rather than deep learning, particularly when students view automated feedback as authoritative. In this sense, Grammarly's pedagogical effectiveness is mediated not only by its technical capabilities but also by students' critical engagement with the feedback it provides.

Taken together, the findings suggest that the pedagogical value of Grammarly lies in its strategic integration rather than its wholesale adoption. From a pedagogical perspective, Grammarly should be positioned as a pre-feedback tool that supports students in addressing surface-level language issues prior to teacher intervention. Such positioning allows instructors to devote greater attention to higher-order writing skills, including content development, argumentation, and discipline-specific discourse conventions. From a student learning perspective, medical students require explicit training in how to use Grammarly critically and reflectively. Rather than accepting automated suggestions uncritically, students should be guided to evaluate, adapt, or reject feedback based on contextual appropriateness and disciplinary accuracy. At the institutional level, universities may consider integrating AWCF tools such as Grammarly into academic writing curricula; however, this integration must be accompanied by systematic teacher training to ensure that such tools are used strategically and pedagogically, rather than as substitutes for meaningful human feedback.

Several limitations of the present study should be acknowledged when interpreting these findings. The relatively small sample size (N = 60) and the short intervention period of 12 weeks may limit the generalizability and robustness of the results, as the sample may not fully represent the diversity of medical students across institutions and regions in Vietnam. Moreover, the duration of the study may not have been sufficient to capture longer-term trajectories of writing development or sustained patterns of engagement with Grammarly. Another limitation concerns the absence of direct data on students' perceptions and experiences. Incorporating learners' voices could provide richer insights into how students interpret automated feedback, the degree of trust they place in it, and how such engagement shapes their learning processes. Building on these limitations, future research should adopt longitudinal designs with larger and more diverse samples, as well as cross-institutional comparisons, to enhance external validity. Further studies could also integrate qualitative methods such as learner interviews, think-aloud protocols, or reflective journals to develop a more nuanced understanding of how Grammarly mediates writing development in medical English contexts.

CONCLUSION

This This study investigated the effectiveness of Grammarly as an Automated Written Corrective Feedback (AWCF) tool in comparison with teacher-provided feedback in supporting medical students' academic writing at Hai Phong University of Medicine and Pharmacy. The findings demonstrate that while both feedback conditions led to statistically significant improvements in writing performance, students who received teacher feedback achieved greater overall gains, particularly in higher-order dimensions such as coherence, argument development, and discipline-specific accuracy. These results underscore the continued centrality of teacher mediation in academic writing instruction, especially within specialized disciplinary contexts.

At the same time, the study confirms that Grammarly offers tangible pedagogical benefits. Its capacity to efficiently identify and correct grammatical, spelling, and lexical errors supported students' independent revision practices and reduced instructors' workload associated with repetitive surface-level corrections. In this sense, Grammarly contributes to enhanced learner autonomy and instructional efficiency. However, the findings also reveal clear limitations, including students' potential over-reliance on automated suggestions and the tool's inability to address complex writing skills such as rhetorical organization, critical engagement with content, and the accurate use of medical terminology.

Taken together, the results suggest that Grammarly is most effective when integrated as a supplementary component of the feedback ecosystem rather than as a replacement for teacher feedback. A blended feedback approach-where Grammarly supports language mechanics and teachers focus on content, structure, and disciplinary conventions-appears to offer the most pedagogically sound model for medical English writing instruction. Such an approach aligns the strengths of automated feedback with the irreplaceable role of human judgment and contextual expertise.

Although the present study is limited by its sample size and duration, it contributes empirical evidence to the relatively

underexplored area of AWCF use in discipline-specific contexts, particularly medical English education in Vietnam. Future research should extend this line of inquiry through larger-scale and longitudinal designs, as well as by incorporating students' perceptions and engagement practices, to deepen understanding of how automated feedback tools shape writing development over time. Overall, while Grammarly can enhance efficiency and support learner autonomy, the findings reaffirm that teacher guidance remains essential for fostering comprehensive academic writing competence in medical education.

REFERENCES

1. Bai, B., & Hu, G. (2017). In the face of fallible feedback: Exploring student responses to teacher written feedback in English writing. System, 65, 83-98. https://doi.org/10.1016/j.system.2016.12.002

2. Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. The JALT CALL Journal, 17(2), 74-92. https://doi.org/10.29140/jaltcall.v17n2.303

3. 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. https://doi.org/10.46451/ijts.2021.06.05

4. Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. Studies in Second Language Acquisition, 32(2), 181-201. https://doi.org/10.1017/S0272263109990490

5. Ghufron, M. A. (2019). Exploring an automated feedback program "Grammarly" and teacher corrective feedback in EFL writing assessment. In Proceedings of the 3rd English Language and Literature International Conference (ELLiC) (pp. 156-161). Universitas Muhammadiyah Semarang.

6. Hyland, K., & Hyland, F. (2019). Feedback in second language writing: Contexts and issues (2nd ed.). Cambridge University Press.

7. Iskender, A. (2023). Holy or unholy? Interview with Open AI's ChatGPT. European Journal of Tourism Research, 34, 3414. https://doi.org/10.54055/ejtr.v34i.3144

8. Li, Z., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. Journal of Second Language Writing, 27, 1-18. https://doi.org/10.1016/j.jslw.2014.10.004

9. Mao, S., & Crosthwaite, P. (2019). Investigating written corrective feedback: (Mis)alignment of teachers' beliefs and practice. Journal of Second Language Writing, 45, 46-60. https://doi.org/10.1016/j.jslw.2019.05.004

10. Nguyen, T. T. (2020). Challenges in academic writing for Vietnamese EFL medical students. Vietnam Journal of Medical Education, 14(3), 25-32.

11. O'Neill, R., & Russell, A. M. T. (2019). Grammarly: Help or hindrance? Academic learning advisors' perceptions of an online grammar checker. Journal of Academic Language and Learning, 13(1), A88-A107.