Research Article

GRAMMARLY AS AUTOMATED WRITTEN CORRECTIVE FEEDBACK IN EFL STUDENTS' RECOUNT WRITING

Ariyani[1*], Mardiana[1], Andi Asmawati[1]
[1]*English Education Department, Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Alauddin Makassar, Gowa, Indonesia*
**Corresponding author: arihyanhi10@gmail.com*

Abstract

Writing is a demanding productive skill for EFL learners, and grammatical accuracy is among their most persistent difficulties. AI-based grammar checkers such as Grammarly provide automated written corrective feedback and are widely used, yet evidence of their effect on the overall quality of university students' writing in Indonesia remains limited. This study examined the effect of Grammarly on the recount writing of English Education Department students at UIN Alauddin Makassar. A pre-experimental one-group pre-test–post-test design was used with 28 third-semester students selected through purposive sampling. Recount texts were scored by two raters with Jacobs et al.'s analytic rubric, and semi-structured interviews were conducted with three students. Because the post-test scores were not normally distributed, the Wilcoxon signed-rank test was used. The mean score rose from 65.00 (SD = 13.03) to 73.57 (SD = 9.32), and the difference was significant with a large effect (Z = −3.35, p = .001, r = .63). All five components improved, most in language use and organization and least in mechanics. The interviewees viewed Grammarly positively, but two of them accepted every suggestion without checking it, whereas the student who evaluated suggestions critically was the only interviewee whose score fell slightly. The findings are discussed in terms of the Noticing Hypothesis and research on automated feedback, with attention to the alternative explanations that a one-group design leaves open. Grammarly appears most useful as a supporting tool within guided writing instruction.

Keywords: *Automated written corrective feedback; Grammarly; Noticing Hypothesis; Recount text; Writing*

Article History:
Submitted : 6 October 2026
Revised : 13 October 2026
Accepted : 14 October 2026
Published : 18 October 2026

How to cite this article:
Ariyani, Mardiana, & Asmawati, A. (2026). Grammarly as automated written corrective feedback in EFL students' recount writing. *ELSTIC-IJ: English Language, Linguistics, and Culture International Journal, 6*(02), https://doi.org/10.24252/elstic-ij.v6i02.73320

Copyright © *2026 by the Authors. Published by ELSTIC-IJ, UPT Pusat Pengembangan Bahasa, Universitas Islam Negeri Alauddin Makassar. This is an open-access article distributed under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0): https://creativecommons.org/licenses/by-sa/4.0/.*

INTRODUCTION

The development of technology in the era of the Industrial Revolution 4.0 has transformed education, including language teaching, by integrating digital tools that improve access, efficiency and the overall learning experience (Hayudiyani et al., 2020; Sorooshian & Panigrahi, 2020). Among the four language skills, writing has been strongly influenced by

this integration. As a productive skill, writing is a key competency that Indonesian EFL students need for academic and professional communication (Toba et al., 2019), supporting academic success and later career prospects when students write scholarship essays, proposals and workplace reports (Moses & Mohamad, 2019). The value of writing is also emphasised in Islamic teachings, as reflected in the oath by the pen in Surah Al-Qalam (68:1).

Despite its importance, many students find writing difficult. Insufficient vocabulary and grammar are among the most significant obstacles (Alisha et al., 2019), alongside problems with syntax, coherence, idea development and organisation (Ahmed, 2019). Indonesian EFL learners in particular struggle with grammar, idea development, vocabulary use and text organisation (Asnas & Hidayanti, 2024; Toba et al., 2019). A preliminary study conducted on 31 May 2025 with second-semester students of the English Education Department at UIN Alauddin Makassar confirmed this picture: most students had difficulty with grammar, vocabulary and coherent organisation when writing texts.

Feedback is one of the main ways in which teachers help students overcome these difficulties, but providing detailed written corrective feedback on every draft is time-consuming, particularly in large classes (Bitchener & Ferris, 2012). AI-based grammar checkers such as Grammarly offer immediate, error-specific feedback and can therefore be understood as a form of automated written corrective feedback (AWCF) (Koltovskaia, 2020; Stevenson & Phakiti, 2014). Previous research suggests that such feedback improves grammatical accuracy, but also that gains in accuracy do not always transfer to overall text quality (Stevenson & Phakiti, 2014) and that students vary greatly in how they engage with automated suggestions (Koltovskaia, 2020). Evidence on whether Grammarly improves the overall quality of Indonesian university students' writing, assessed across content, organisation, vocabulary, language use and mechanics, remains limited. This study addresses three research questions:

RQ1. What is the quality of English Education Department students' recount writing before instruction using Grammarly?

RQ2. What is the quality of their recount writing after instruction using Grammarly?

RQ3. Is there a significant difference between students' recount writing before and after instruction using Grammarly?

LITERATURE REVIEW

Writing Quality and Recount Text

In this study, writing quality refers to students' ability to produce a coherent, well-organised and grammatically accurate recount text, covering content, organisation, vocabulary, language use and mechanics. A recount text reconstructs past events or experiences to inform or engage readers (Anderson & Anderson, 1997, as cited in Harris et al., 2014), so it requires chronological organisation as well as linguistic accuracy. Producing one is challenging because of limited vocabulary, weak grammar, difficulty connecting ideas, poor control of mechanics, and limited exposure to English and authentic writing practice (Moses & Mohamad, 2019; Phi et al., 2021). These challenges make tools that provide immediate linguistic feedback particularly relevant.

Technology and Automated Feedback in Language Learning

Technology has long been part of language learning, from audio, video and whiteboards to computer-based and mobile resources connected to the Internet (Chun et al., 2016). Computer-assisted language learning (CALL) allowed learners to study independently and practise language skills through interactive software (Beatty, 2010), and the spread of smartphones gave rise to mobile-assisted language learning (MALL) (Jarvis & Achilleos, 2013), which fosters self-regulated learning by letting learners adjust the pace and content of instruction and access materials anytime and anywhere (Puebla et al., 2022). In such environments teachers increasingly act as facilitators, while students are expected to manage their own learning (Dhamayanti & Farah, 2021).

Written corrective feedback (WCF) has been studied extensively in second language writing, and there is broad agreement that it can improve learners' accuracy, particularly when learners attend to and act on it (Bitchener & Ferris, 2012). Automated written corrective feedback extends this tradition by generating feedback through software. In a review of studies on computer-generated feedback, Stevenson and Phakiti (2014) found modest evidence that such feedback improves the quality of the texts on which it is given, but much weaker evidence that it improves writing proficiency more generally, and they noted that gains in accuracy do not always transfer to overall text quality. Koltovskaia (2020) showed that students engage with Grammarly's feedback in very different ways: some evaluate suggestions critically, while others accept them with little cognitive engagement, which may limit learning.

Grammarly as an AI Grammar Checker

Artificial intelligence (AI) has become a prominent part of language learning because of its capacity to deliver immediate feedback and personalised learning opportunities (Kuddus, 2022). Among AI writing tools, Grammarly and QuillBot are reported to be the most frequently used by students for academic English writing (Zulfa et al., 2023). QuillBot is designed mainly for paraphrasing, whereas Grammarly focuses on grammar, spelling, punctuation and style (Ma'ruf, 2025). Grammarly detects problems in contextual spelling, grammar, punctuation, sentence structure, style and vocabulary, and suggests revisions (Fitriana & Nurazni, 2022). It can be used through web browsers, desktop software, smartphones and Microsoft Word. Basic checks are available in the free version, while feedback on clarity, conciseness, tone and plagiarism belongs to the premium version. By combining natural language processing and machine learning, Grammarly gives immediate feedback that can support self-regulation and learner independence (Turdaliyevna, 2024), and it is regarded as one of the most comprehensive grammar checkers available (Ismailia & Novawan, 2025).

Theoretical Perspectives

Two theories frame this study. Schmidt's (1990, 2001) Noticing Hypothesis holds that exposure to input is not enough for learning; learners must consciously attend to linguistic features for input to become intake. In writing, students must become aware of their grammatical, lexical and mechanical errors before they can correct them, and attention to form facilitates the internalisation of correct structures (Ellis, 2008; Gass & Selinker, 2008). Grammarly may promote noticing by highlighting errors and offering revisions, drawing attention to the gap between learners' current language use and the target form. Noticing in Schmidt's sense, however, involves the conscious registration of a form. A learner who clicks to accept a suggestion without reading it may change the text without noticing anything, so automated feedback can either promote or bypass noticing depending on how it is used. Mayer's (2001) Cognitive Theory of Multimedia Learning suggests why such feedback may be processed effectively when it is attended to: learning is enhanced when information is delivered through both visual and verbal channels, and Grammarly combines colour-coded underlines with written explanations.

Previous Studies

Previous studies generally report positive effects of Grammarly. Huang et al. (2020) found that the mean writing score of Chinese university EFL students rose from 58.05 to 73.90 over 16 weeks ($p < .001$), although the tool worked best together with peer and teacher feedback. Vu (2024) reported a noticeable decline in grammatical errors among 32 Vietnamese college students. Dizon and Gayed (2021) found that Grammarly significantly improved the grammatical accuracy and lexical variation of 31 Japanese university students' mobile writing, but not their syntactic complexity or fluency. In Indonesia, Jelita et al. (2023) used a one-group pre-test–post-test design with 30 senior high school students and found fewer errors in grammar, spelling, capitalisation and punctuation, while Hadiat (2022) reported improved accuracy in descriptive writing among junior high school students. More recent studies also associate Grammarly with better writing performance (Estrella, 2025; Hidayat & Ratna, 2025; Rababah & Talafha, 2024).

Most of these studies involved university students abroad or secondary school students in Indonesia, and they mainly examined grammatical accuracy, perceptions or specific error types. The present study differs from earlier work in three ways. It involves university students majoring in English Education in Indonesia; it assesses writing through five analytic components rather than grammatical accuracy alone, which allows the transfer question raised by Stevenson and Phakiti (2014) to be examined; and it uses recount text as the writing task, whereas earlier studies examined essays, explanatory texts or descriptive texts. It also combines test scores with interviews about how students used Grammarly's suggestions.

METHOD

Research Design

This study applied a pre-experimental one-group pre-test–post-test design, in which a single group is measured before and after a treatment (Gay et al., 2011). Brief semi-structured interviews were added to help interpret the quantitative findings (Creswell & Creswell, 2017). The study was conducted in the English Education Department, Faculty of Tarbiyah and Teacher Training, UIN Alauddin Makassar, in the 2025/2026 academic year, and its scope was limited to recount writing by third-semester students.

Participants

The population comprised 115 third-semester students of the class of 2024, divided into four classes (PBI A, B, C and D). The sample was selected through purposive sampling based on four criteria: (a) being a third-semester student of the department, (b) having completed or currently taking at least one basic writing course, (c) being familiar with recount text, and (d) owning or having access to a smartphone or laptop that can run Grammarly. Twenty-eight students from one class, PBI C, met these criteria and participated. Three of them were later interviewed and are referred to as R1, R2 and R3.

Instruments

The main instrument was a writing test in which students wrote a recount text of approximately 150–180 words. Writing tests provide measurable evidence of how well students organise ideas, use grammar and produce coherent texts (Creswell & Creswell, 2017; Masrul & Erliana, 2024). The texts were scored with the analytic rubric of Jacobs et al. (1981), which distributes a maximum of 100 points across five components (Table 1) and describes each component at four performance levels.

Table 1. Components of the writing rubric adapted from Jacobs et al. (1981)

Component	Maximum score	Aspects assessed
Content	30	Understanding of the topic, amount and relevance of information, development of the main idea
Organization	20	Clarity of expression, connection of supporting details to main ideas, logical sequence, coherence
Vocabulary	20	Range of vocabulary, accuracy of word choice and word forms, appropriateness to the context
Language use	25	Sentence construction; agreement, tense, word order, articles, pronouns and prepositions
Mechanics	5	Spelling, punctuation, capitalisation and paragraphing
Total	**100**	

Two raters, the first author and another student of the department who had passed the writing and language assessment courses, scored all pre-test and post-test texts independently using the same rubric. The test was not formally validated through expert judgement or pilot testing. A semi-structured interview with six questions, adapted from Maypida and Damanik (2024) and Nurlia et al. (2025), explored participants' general perceptions, the most improved aspect, their revision process, ease of use, difficulties and willingness to continue using Grammarly.

Procedure

The study consisted of eight meetings over 8 weeks. In the first meeting, students wrote the pre-test recount text on the topic given within 90 minutes, without any digital tools. The treatment took place in meetings two to seven (Table 2). Students used the premium version of Grammarly on their browsers, smartphones or laptops, and were shown how to interpret its suggestions on correctness and clarity. Over the treatment, each student wrote x recount texts and revised x drafts with Grammarly. The lecturer also gave feedback on drafts, and how participation was monitored (e.g., screenshots of Grammarly reports, submission of first and revised drafts).

Table 2. Treatment meetings

Meeting	Aim	Activities	Duration
2	Introduce Grammarly	Account set-up; tour of correctness and clarity suggestions	90 min
3	Understand Grammarly's feedback	Analysis of a model recount text and of Grammarly's suggestions on it	90 min
4	Guided writing practice	Joint construction and guided writing of a recount text	90 min
5	Independent drafting	Writing a first full draft of a recount text	90 min
6	Revision with Grammarly	Revising the draft using Grammarly's suggestions	90 min
7	Reflection	Reflecting on accepted and rejected suggestions and recurring errors	90 min

In the eighth meeting, students wrote the post-test recount text on a different topic from the pre-test, within 90 minutes and without access to Grammarly or any other digital tool, so that the post-test reflected students' own writing rather than tool-corrected text. After the post-test, three students were selected for interviews because they had fully participated in the treatment, were willing to be interviewed and represented different levels of writing performance. The interviews were held in June 2026, two face to face and one online, in Indonesian, and lasted about x minutes each.

Data Analysis

Total scores were classified using the categories of Jacobs et al. (1981) shown in Table 3, and the percentage of students in each category was calculated. Data were analysed in IBM SPSS Statistics 26. Descriptive statistics (mean, standard deviation, median, interquartile range, minimum and maximum) were computed for both tests. Because the normality assumption of the paired t-test concerns the difference scores, the Shapiro–Wilk

test was applied to the post-test minus pre-test differences. As the post-test scores were not normally distributed and the difference scores were not normally distributed, the Wilcoxon signed-rank test was used as the main test, and a paired-samples t-test was run as a robustness check. The effect size was calculated as $r = Z/\sqrt{N}$, where N is the number of pairs. The significance level was .05.

Table 3. Classification of students' scores (Jacobs et al., 1981)

Score range	Proficiency evaluation
88–100	Very good
75–87	Good
64–74	Fair
49–63	Poor
34–48	Very poor

The interviews were audio-recorded, transcribed and analysed thematically (Braun & Clarke, 2006). Transcripts were coded deductively according to the six interview topics, and codes were grouped into three themes: perceptions and improved aspects, revision strategies, and difficulties and continued use. Quotations were translated from Indonesian into English by the first author and checked by the second author.

FINDINGS AND DISCUSSION

Students' Recount Writing Before the Treatment (RQ1)

Table 4 presents the classification of the pre-test scores. Before using Grammarly, the largest group of students was in the poor category (11 students, 39.29%), only one student (3.57%) reached the very good category, and three students (10.71%) were in the very poor category. In total, 22 of the 28 students (78.57%) scored in the fair category or below. The mean pre-test score was 65.00 (SD = 13.03), with scores ranging from 35 to 91 (Table 6), showing that the students' ability varied considerably. This is in line with earlier studies reporting that Indonesian EFL learners struggle to control grammar, develop ideas, choose vocabulary and organise texts (Asnas & Hidayanti, 2024; Toba et al., 2019).

Table 4. Classification of students' pre-test and post-test scores

Classification	Score	Pre-test f	Pre-test %	Post-test f	Post-test %
Very good	88–100	1	3.57	0	0.00
Good	75–87	5	17.86	16	57.14
Fair	64–74	8	28.57	8	28.57
Poor	49–63	11	39.29	4	14.29
Very poor	34–48	3	10.71	0	0.00
Total		**28**	**100.00**	**28**	**100.00**

Students' Recount Writing After the Treatment (RQ2)

After the treatment, the distribution of scores shifted upward (Table 4). The good category became the largest, with 16 students (57.14%), the very poor category disappeared, and the poor category fell from 39.29% to 14.29%. The fair category remained at 28.57%. No student reached the very good category in the post-test, although one student had done so in the pre-test. The mean score increased from 65.00 to 73.57 and the standard deviation decreased from 13.03 to 9.32, so the spread of scores narrowed (Table 6). This narrowing should be interpreted with caution. In a one-group design, students with very low pre-test scores tend to score higher on retest and those with very high scores lower, so part of the reduction may reflect regression to the mean rather than equal support from Grammarly for students at different levels (Shadish et al., 2002). The mean scores of both tests fall within the fair category, although the post-test mean is close to the threshold of the good category, and the proportion of students in the poor or very poor categories fell from 50.00% to 14.29%.

Improvement in the writing components

Table 5 and Figure 1 compare the mean score of each component before and after the treatment. All five components improved. In raw points, language use showed the largest gain (2.61), followed by organization (2.43), content (1.75) and vocabulary (1.28), while mechanics showed the smallest gain (0.50). Because the components carry different maximum scores, raw gains are not directly comparable: as a percentage of each component's maximum, organization gained most (12.2%), followed by language use (10.4%) and mechanics (10.0%), while content (5.8%) and vocabulary (6.4%) gained least. The component gains are descriptive and were not tested separately for significance.

Table 5. Mean scores of the writing components and gains

Component	Maximum	Pre-test	Post-test	Gain	Gain (% of max.)
Content	30	21.39	23.14	1.75	5.8
Organization	20	13.92	16.35	2.43	12.2
Vocabulary	20	13.00	14.28	1.28	6.4
Language use	25	13.78	16.39	2.61	10.4
Mechanics	5	2.89	3.39	0.50	10.0

Note. Gain = post-test mean minus pre-test mean.

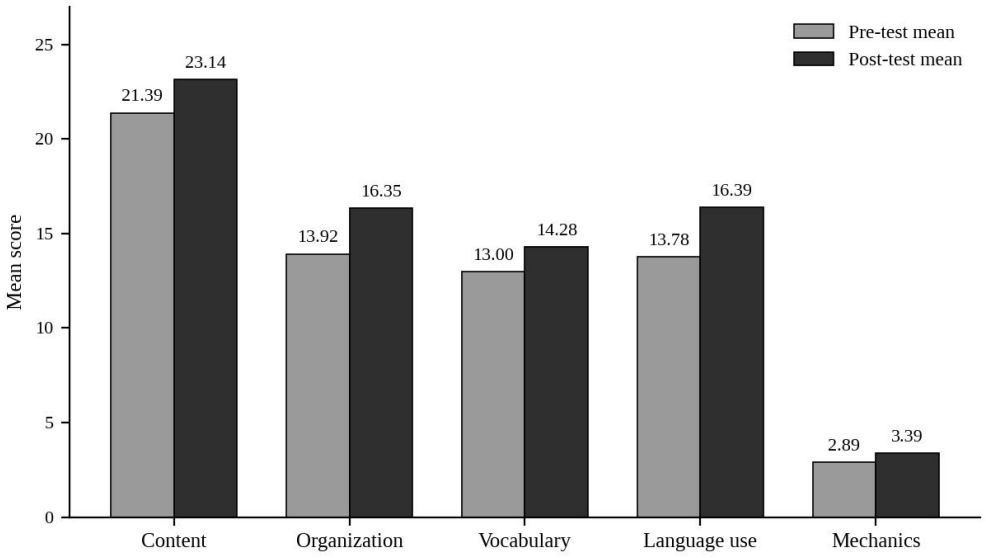


Figure 1. Mean scores of the writing components in the pre-test and post-test

The largest gain in language use corresponds to Grammarly's main function as a grammar checker, which highlights errors such as subject–verb agreement, verb tense and articles and explains each correction. Language use also had the lowest starting point relative to its maximum (55.1%, compared with 71.3% for content and 69.6% for organization), so it had the most room to improve. The gain in organization requires a more careful explanation. Grammarly works mainly at sentence level; its clarity suggestions may improve the flow of individual sentences, but it does not evaluate paragraph structure or chronological sequencing. The organization gain may therefore owe more to the treatment meetings themselves, which included analysis of a model recount text, guided writing practice, drafting and reflection, than to the tool. Vocabulary improved moderately, possibly because Grammarly suggests synonyms for repeated words, and content improved least because Grammarly cannot judge the depth or relevance of ideas. Mechanics improved least in raw points despite Grammarly's ease in detecting spelling, punctuation and capitalisation errors. With a maximum of only 5 points and a pre-test mean of 2.89, this component offers little room to show change, so a

more fine-grained measure, such as mechanical errors per 100 words, would capture mechanical accuracy better in future studies.

The Effect of Grammarly on Students' Recount Writing (RQ3)

Table 6 presents the descriptive and inferential statistics. The Shapiro–Wilk test showed that the pre-test scores were normally distributed (W = .962, p = .394) but the post-test scores were not (W = .896, p = .009). The Wilcoxon signed-rank test showed that post-test scores were significantly higher than pre-test scores, Z = −3.35, p = .001, with a large effect size (r = .63).

Table 6. Descriptive statistics and Wilcoxon signed-rank test

Statistic	Pre-test	Post-test
N	28	28
Mean (SD)	65.00 (13.03)	73.57 (9.32)
Median (IQR)	65	75
Minimum–maximum	35–91	53–86
Shapiro–Wilk W, p	.962, .394	.896, .009
Wilcoxon Z, p	−3.35, .001	
Effect size r	.63	

Note. r = Z/√N, with N = 28 pairs. The Kolmogorov–Smirnov results were D = .142, p = .158 (pre-test) and D = .183, p = .017 (post-test).

These results can be interpreted through Schmidt's (1990) Noticing Hypothesis. When Grammarly underlines errors and offers revisions, students can compare their original sentences with the suggested forms and become aware of the gap between what they wrote and the target language, which may turn feedback into intake. Mayer's (2001) multimedia principle offers a complementary explanation, since Grammarly presents feedback both visually and verbally. These are theoretical interpretations, however; the study did not measure noticing or cognitive processing directly. The finding that all five components improved, and not only language use, also bears on the transfer question raised by Stevenson and Phakiti (2014). Because the post-test was written without Grammarly, the gains suggest some carry-over into students' unaided writing, although, as noted above, part of the improvement in organization and content may reflect the instruction that accompanied the tool rather than the tool itself.

The finding is consistent with previous studies reporting that Grammarly improves students' writing (Dizon & Gayed, 2021; Estrella, 2025; Hadiat, 2022; Hidayat & Ratna, 2025; Rababah & Talafha, 2024; Vu, 2024) and with Jelita et al. (2023), who used the same design

and statistical test with senior high school students. The gain was smaller than in Huang et al. (2020), whose students improved by 15.85 points over 16 weeks compared with 8.57 points over four weeks here, but the studies differ in participants, tasks and scoring, so only the direction of the effect can be compared. Like Dizon and Gayed (2021), who found no improvement in syntactic complexity or fluency, the present study suggests that Grammarly is more effective for accuracy-related aspects than for developing ideas. A significant Wilcoxon result shows that post-test scores differ systematically from pre-test scores, but in the absence of a control group it does not establish that Grammarly caused the difference.

Students' Responses to Grammarly

Table 7 summarises the responses of the three interviewees. R1 was a high-performing student, while R2 and R3 represented middle and lower levels of performance.

Table 7. Summary of interview responses

Aspect	R1	R2	R3
General perception	Positive; helped grammar accuracy and made writing easier	Positive; corrected her disorganised grammar; more reliable than other tools	Positive; helpful for errors common among L2 learners
Most improved aspect	Organization and vocabulary	Vocabulary (synonyms, precise word choice)	Organization (subject placement, sentence order), then vocabulary
Use of suggestions	Selective; checked each suggestion	Accepted all suggestions	Accepted all suggestions
Difficulties	None	Login problem; Android device incompatible	Login problem at first
Future use	Probably; interested in other apps	Yes; most enthusiastic	Yes, as a supporting tool only

Perceptions and improved aspects

All three participants viewed Grammarly positively regardless of their performance level. R2 said that her grammar had previously been disorganised and that Grammarly showed which parts needed improvement. R3 found the colour-coded system helpful in identifying the type and location of errors and said that it reduced her uncertainty about whether a sentence was correct. Asked which aspects improved most, R1 named organization and vocabulary, R2 named vocabulary because Grammarly suggested word choices that had not occurred to her, and R3 named organization and then vocabulary. These perceptions match the score gains in organization, although vocabulary showed a smaller quantitative gain.

Revision strategies

The participants used the suggestions in different ways. R1 evaluated each suggestion critically because she knew that Grammarly is not always accurate: "the grammarly needs to be be checked because sometimes it is not acuurate as what I mean". R2 and R3, in contrast, accepted every suggestion because they trusted it to follow standard English. This contrast connects directly to the Noticing Hypothesis. Accepting a correction with one click may change the text without the learner consciously registering the form, so it may bypass rather than promote noticing (Schmidt, 2001), and heavy reliance on automated suggestions can increase dependency (Koltovskaia, 2020). Interestingly, R1, the most critical user, was the only interviewee whose post-test score decreased slightly. Critical evaluation may lead students to reject some correct suggestions as well as incorrect ones, and the benefits of more autonomous use, such as better understanding of rules, may not appear in short-term scores. Because only three students were interviewed, these patterns cannot be generalised to the whole class.

Difficulties and continued use

The main difficulty, reported by R2 and R3, concerned logging in. R2 could not use Grammarly on her Android device and used another phone instead, while R3 found the login process difficult at first but had no problems afterwards. None of the participants had trouble with the main features, so the obstacles were technical rather than pedagogical. All three intended to keep using Grammarly. R2 was the most enthusiastic, as this was her first regular use of a grammar checker after relying on translation tools, while R3 gave the most reflective answer: "Grammarly should be a supporting tool, not the main source …" . This view reflects an independent approach to AI-assisted writing in which technology supports the writer's own judgement.

Implications

The results suggest that the components that gained most were those for which Grammarly gives explicit, error-specific feedback, while the interviews show that whether this feedback leads to noticing depends on how students use it. Pedagogically, Grammarly can help lecturers provide more efficient feedback on linguistic accuracy (Huang et al., 2020), allowing them to concentrate on aspects the tool cannot judge, such as the development and organisation of ideas. Its use should, however, be guided: lecturers can ask students to accept or reject each suggestion deliberately and to record and explain recurring errors, turning

automated corrections into opportunities for noticing. Checking device compatibility and account access before instruction begins may also prevent the login problems reported by two participants.

Limitations

The one-group design had no control group, so the improvement cannot be attributed to Grammarly alone; classroom instruction, practice effects and regression to the mean may have contributed. The sample was small and drawn from one class, the writing test was not formally validated, and the second rater was a student rather than a professional assessor. Only three students were interviewed, the treatment lasted four weeks, and the premium version of Grammarly was used, so results may differ with the free version. The findings are therefore best regarded as preliminary evidence.

CONCLUSION

This study examined the effect of Grammarly, used as automated written corrective feedback, on the recount writing of English Education Department students. Before the treatment, students' writing was at a modest level, with half of the students in the poor or very poor categories. After the treatment, the mean score rose from 65.00 to 73.57, and the Wilcoxon signed-rank test showed a significant improvement with a large effect ($Z = -3.35$, $p = .001$, $r = .63$). All five writing components improved, especially language use and organization, and the interviewees perceived Grammarly positively, although two of them accepted all suggestions without checking them.

These findings suggest that Grammarly can support EFL students' writing when it is embedded in guided instruction, but that its contribution to noticing depends on whether students evaluate suggestions rather than accept them automatically. Lecturers are encouraged to integrate Grammarly as supplementary support with tasks that require students to justify their revisions, and students are advised to use it as a learning aid rather than a substitute for understanding grammar. Future research should use a quasi-experimental design with one of the other classes (PBI A, B or D) as a comparison group, include a delayed post-test to examine retention, use finer-grained measures of accuracy such as errors per 100 words, and trace how students engage with individual suggestions.

AI USE AND DISCLOSURE

The authors used generative AI tools Grammarly, quilbot to check grammar, proofread the manuscript, review its clarity and consistency, and assist in searching for relevant references. All references suggested by the tools were checked against the original sources. The authors reviewed and edited all content and take full responsibility for the article.

ETHICAL STATEMENT AND INFORMED CONSENT

This study was approved by the English Education Department, UIN Alauddin Makassar. All participants were informed about the purpose of the study and gave their consent. Participation was voluntary, did not affect course grades, and all scores and interview data were anonymised.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

- Ariyani: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft.
- Mardiana: Supervision, Methodology, Writing – review & editing.
- Andi Asmawati: Supervision, Validation, Writing – review & editing.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this article, and no affiliation with Grammarly Inc.

DATA AVAILABILITY

The anonymised scores and interview transcripts are available from the corresponding author on reasonable request.

ACKNOWLEDGMENTS

The authors thank the students of PBI C, English Education Department, UIN Alauddin Makassar, for taking part in this study.

REFERENCES

Ahmed, P. H. (2019). Major writing challenges experienced by EFL learners in Soran University. *Journal of University of Human Development, 5*(3), 120–126. https://doi.org/10.21928/juhd.v5n3y2019.pp120-126

Alisha, F., Safitri, N., & Santoso, I. (2019). Students' difficulties in writing EFL. *Professional Journal of English Education, 2*(1), 20–25.

Asnas, S. A. M., & Hidayanti, I. (2024). Uncovering EFL students' frequent difficulties in academic writing and the coping strategies: The case of a college in Indonesia. *Journal on English as a Foreign Language, 14*(1), 124–151. https://doi.org/10.23971/jefl.v14i1.7472

Beatty, K. (2010). *Teaching and researching computer-assisted language learning* (2nd ed.). Routledge.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chun, D., Smith, B., & Kern, R. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal, 100*(S1), 64–80. https://doi.org/10.1111/modl.12302

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Dhamayanti, I., & Farah. (2021). EFL students' perception and motivation toward Quizizz as e-learning media in English e-classroom. *Education of English as Foreign Language, 4*(2), 71–78. https://doi.org/10.21776/ub.educafl.2021.004.02.03

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.v17n2.336

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.

Estrella, F. (2025). Supporting independent writers: How Grammarly shapes learner autonomy and writing performance in the Ecuadorian foreign language context. *Social Education Research, 6*(1), 125–137.

Fitriana, K., & Nurazni, L. (2022). Exploring students' perception of using Grammarly to check grammar in their writing. *JET (Journal of English Teaching), 8*(1), 15–25. https://doi.org/10.33541/jet.v8i1.3044

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). Taylor & Francis.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Educational research: Competencies for analysis and applications* (10th ed.). Pearson Education.

Hadiat, A. W. F. (2022). The use of Grammarly to enhance students' accuracy in writing descriptive text (A case study at eighth grade of a junior high school in Ciamis). *Journal of English Education Program (JEEP), 9*(2), 133. https://doi.org/10.25157/(jeep).v9i2.8552

Harris, A., Ansyar, M., & Radjab, D. (2014). An analysis of students' difficulties in writing recount text at tenth grade of SMA N 1 Sungai Limau. *Journal English Language Teaching (ELT), 2*, 55–63.

Hayudiyani, M., Mustiningsih, & Arifin, I. (2020). Reorientation of curriculum in the face of Industrial Revolution 4.0. *Advances in Social Science, Education and Humanities Research, 508*, 659–664. https://doi.org/10.2991/assehr.k.201214.314

Hidayat, A., & Ratna, S. (2025). The effect of using Grammarly on EFL students' writing accuracy: A quasi-experimental research. *Journal of Linguistics and Social Sciences, 3*(2), 75–82.

Huang, H. W., Li, Z., & Taylor, L. (2020). The effectiveness of using Grammarly to improve students' writing skills. In *Proceedings of the 5th International Conference on Distance Education and Learning* (pp. 122–127). ACM. https://doi.org/10.1145/3402569.340259

Ismailia, T., & Novawan, A. (2025). Enhancing EFL students' recount writing: A comparative analysis of Grammarly, QuillBot, and GrammarCheck applications. *Journal of English in Academic and Professional Communication, 11*(2), 89–102. https://doi.org/10.25047/jeapco.v11i2.5095

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House.

Jarvis, H., & Achilleos, M. (2013). From computer assisted language learning (CALL) to mobile assisted language use (MALU). *TESL-EJ, 16*(4), 1–18.

Jelita, K. N., Daud, A., & Masyhur, M. (2023). The effectiveness of using Grammarly on high school students' writing quality. *International Journal of Educational Best Practices, 7*(1), 43–55. https://doi.org/10.31258/ijebp.v7n1.p43-55

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*, Article 100450. https://doi.org/10.1016/j.asw.2020.100450

Kuddus, K. (2022). Artificial intelligence in language learning: Practices and prospects. In *Advanced analytics and deep learning models* (pp. 3–17). Wiley. https://doi.org/10.1002/9781119792437.ch1

Ma'ruf, M. I. (2025). *QuillBot dan Grammarly: Kolaborasi AI untuk tulisan akademik*. Indonesia Artificial Intelligence Hub. https://aihub.id/opini/quillbot-grammarly-ai-tulisan-akademik

Masrul, M., & Erliana, S. (2024). Students' writing test: An argumentative study of English as a foreign language learner. *Journal of Education and Learning, 18*(3), 753–761. https://doi.org/10.11591/edulearn.v18i3.20971

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Maypida, A., & Damanik, E. S. D. (2024). Exploring Grammarly's effectiveness in EFL writing through a student perception study. *Journal of Educational Development, 12*(2), 98–107.

Moses, R. N., & Mohamad, M. (2019). Challenges faced by students and teachers on writing skills in ESL contexts: A literature review. *Creative Education, 10*(13), 3385–3391. https://doi.org/10.4236/ce.2019.1013260

Nurlia, I., Eliwarti, & Supriusman. (2025). Students' perceptions of using Grammarly application in enhancing their writing ability. *Journal of English Language Pedagogy, Literature, and Culture, [[volume]]*(3), 333–350.

Phi, V., Pham, H., Kim, T., & Bui, L. (2021). Genre-based approach to writing in EFL contexts. *World Journal of English Language, 11*(2), 95–106. https://doi.org/10.5430/wjel.v11n2p95

Puebla, C., Fievet, T., Tsopanidi, M., & Clahsen, H. (2022). Mobile-assisted language learning in older adults: Chances and challenges. *ReCALL, 34*(2), 169–184. https://doi.org/10.1017/S0958344021000276

Rababah, L. M., & Talafha, D. K. (2024). Unlocking writing potential: Assessing the impact of Grammarly on Jordanian EFL students' writing proficiency. *Theory and Practice in Language Studies, 14*(5), 1486–1492. https://doi.org/10.17507/tpls.1405.21

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Sorooshian, S., & Panigrahi, S. (2020). Impacts of the 4th Industrial Revolution on industries. *Walailak Journal of Science and Technology, 17*(8), 903–915.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Toba, R., Noor, W. N., & Sanu, L. O. (2019). The current issues of Indonesian EFL students' writing comparison and contrast essay. *Dinamika Ilmu, 19*(1), 57–73.

Turdaliyevna, B. B. (2024). Using artificial intelligence technologies in language teaching. *International Journal of Literature and Languages, 4*(11), 35–39. https://doi.org/10.37547/ijll/Volume04Issue11-08

Vu, P. T. (2024). Using Grammarly in enhancing students' grammatical accuracy in English writing: A case at a private college in Hanoi. *Proceedings of the AsiaCALL International Conference, 6*, 81–96. https://doi.org/10.54855/paic.2466

Zulfa, S., Dewi, R. S., Hidayat, D. N., Hamid, F., & Defianty, M. (2023). The use of AI and technology tools in developing students' English academic writing skills. *Proceedings of the Annual International Conference on Education, 1*, 47–63.