



PERCEPTIONS OF *GRAMMARLY*'S FREE VERSION AMONG INDONESIAN EFL UNDERGRADUATES: EVIDENCE FROM ARGUMENTATIVE ESSAY WRITING

Hanna Maryam Azzahra¹, Lilik Ulfiati², Duti Volya³

¹hannamaryamazzahra32@gmail.com, ²lilik.ulfiati@unja.ac.id, ³dutivolya@unja.ac.id

¹Universitas Jambi, Muaro Jambi, Jambi, Indonesia

²Universitas Jambi, Muaro Jambi, Jambi, Indonesia

³Universitas Jambi, Muaro Jambi, Jambi, Indonesia

ABSTRACT

This study examined Indonesian sixth-semester EFL undergraduates' perceptions of *Grammarly*'s free version and of their own argumentative essay quality, along with the relationship between the two. A quantitative, cross-sectional descriptive-correlational design was applied, involving 107 learners from a public university in Jambi Province selected through purposive sampling. Data were collected via a 28-item questionnaire assessing four *Grammarly* characteristics (spell-checker, grammar checker, punctuation, clarity) and seven argumentative-essay components which were analyzed by descriptive statistics, Pearson correlation, and simple linear regression. Results showed learners had positive perceptions of *Grammarly*, nonetheless those perceptions varied by feature ($M = 3.06$, high), with clarity rated highest and spell-checker lowest. Perceptions of essay quality were also favorable ($M = 3.07$, high), with concluding paragraphs rated most positively and the introduction the lowest. A moderately strong, positive relationship emerged between the two variables ($R^2 \approx .41$, $r \approx .64$), demonstrating that learners who perceived *Grammarly* more favorably also rated their essay quality more highly. These findings recommend *Grammarly* free version is perceived as most useful for refining and summarizing already-formed language, whereas offering limited support for higher-order tasks namely background contextualization and argument development. The study implies *Grammarly* should complement instructors' feedback in EFL argumentative instruction.

Keywords: Argumentative Essay, Automated Writing Evaluation, EFL Writing, *Grammarly*, Learner Perceptions

A. INTRODUCTION

Writing an argumentative essay is widely concerned as one of the most challenging tasks that undergraduate English as Foreign Language (EFL) students have to learn, since it needs either grammatical accuracy or ability to make a clear thesis, develop coherent supporting paragraphs, anticipate counterarguments, and end with logical conclusion (Hyland, 2019). This difficulty is frequently experienced by the learners in writing argumentative texts across EFL settings worldwide. Ghanbari & Salari (2022) found that Iranian EFL undergraduate were mostly unable to critically assess their own evidence or to construct integrative, coherent-arguments, and their instructors supported the idea that the issue came from underdeveloped critical-thinking preparation, not only from language proficiency. Learners' cultural and linguistic diversities in some other countries also cause their problems in writing the argumentative texts.

Another research conducted by Ozfidan & Mitchell (2020) reported that Saudi-background EFL students struggled to organize claims in a logical way and support them with sufficient evidence which indicates that argument structure trouble experienced by across different EFL population. In Indonesia, moreover, Ananda et al., (2024) found that EFL undergraduates at a public university in Jambi Province had difficulties to produce and arrange arguments, construct a clear thesis statement, and integrate counterclaims and refutations. Similarly, the students were lack of strategies to find and incorporate supporting evidence. These results are in line with the previous studies documented internationally, and they show very pressing need for context-sensitive writing support in this region.

For EFL learners like in Indonesia, they encountered some causes to write the argumentative texts such as limited exposure to authentic English input, lack of chances to do a lot of writing practices, continuing trouble at the sentence level, and text organization. Pasaribu et al., (2024) found that Indonesian EFL undergraduates' academic writing was constrained by restricted experience to real English texts, grammatical and organizational weaknesses. Problems with discourse-level organization are not merely an Indonesia thing either, Xie and Lv (2022) indicated that without structured content or planning assistance, EFL undergraduates' argumentative essays often missed proper discourse links between ideas. Compounding these linguistic pressures, chances for sustained, individualized feedback within the classroom itself are structurally limited, a wide-ranging synthesis across decade by Yu and Yang (2021) showed that the amount and consistency of teacher feedback ESL/ EFL learners get is frequently constrained by classroom size and workload, thus timely response to student writing becomes difficult. Computer-assisted language learning tools, therefore, have been supported as means to facilitate L2 writers by offering well-timed individualized feedback which regular classroom teaching might not be able to provide, particularly when the writers have to deal with the scale (Warschauer & Healey, 1998). More recent scholarship confirms this affordance that consider automated written corrective feedback has been used to assist students rework mistakes on their own, regardless of whether the teacher is around, (Ranalli, 2018) and automated feedback delivered through computer-assisted platform has supported the development of L2 written causal explanations (Saricaoglu, 2019). Furthermore, Hassanzadeh and Fotoohnejad (2021) highlighted that a classroom integrated automated feedback program let learners keep getting, diagnostic, individualized replies about their writing, on a continued basis, which seemed to cover for the feedback gaps that often appear under teacher centered instruction. From a sociocultural lens, Xu and Zhang (2022) argued that automated writing evaluation feedback is going to interact with learners' proficiency levels to mediate their revision processes so these tools have functions as a scalable add on which are not a total substitution for teacher feedback in resource constrained EFL classrooms.

Among the many digital feedback tools available nowadays, automated writing evaluation (AWE) system have been given attention in EFL writing research because they provide learners with instantaneous tool-generated feedback on surface-level and discourse level features of their writing. *Grammarly*, a freely accessible AWE application has been popular tools used by EFL undergraduates for checking spelling, grammar, punctuation, and clarity in academic writing tasks. Previous studies concerning on *Grammarly*'s effectiveness, yet, remain mixed and feature-dependent rather than uniformly positive. Dewi (2022) noticed that EFL students perceived *Grammarly*'s automated evaluation as effective mainly for correctness related parts of their writing, whereas its usefulness for higher order for instance argument development was less helpful. Similarly, Armanda et al., (2022) also found that EFL students rated *Grammarly*

beneficially for Grammar and mechanics but tended more reserved views regarding its contribution to stylistic sophistication. Furthermore, in studies by Samosir and Daulay (2023) and Komara and Putra (2024) were reported that feature-specific instead of globally uniform learner perceptions of *Grammarly* as an AWE tool, and by Inayah and Apoko (2024) documented that EFL students have positive views in writing analytical exposition texts by guided *Grammarly*'s grammar but free version *Grammarly* has limited support for content and structural development,

A closely related line of research has studied how *Grammarly*'s feedback links to the components of argumentative and academic essay writing. Hoang and Van (2025) investigated the precision of *Grammarly*'s feedback on EFL learners' writing and identified that students perceived stronger support for surface-level correctness than for the development of argumentation. Moreover, Thi and Nikolov (2022) reported that *Grammarly* feedback worked best alongside teacher feedback for language accuracy, while teacher input was still needed for organization and argumentation. This distinction is kind of track ongoing worries about whether AWE tools can significantly assist the higher-order, rhetorical aspects of argumentative writing, namely thesis clarity, counterargument and refutation, as well as logical coherence. Beyond perception studies, more intervention research suggests that Artificial Intelligent (AI)-assisted feedback, including *Grammarly*-based feedback, can bring measurable improvements in writing quality. Alnemrat et al., (2025) reported that AI-generated feedback led to writing-performance gains which were comparable to teacher feedback on an argumentative writing task among EFL undergraduates, although Barrot (2023) found that automated written corrective feedback enhanced L2 writing accuracy. Taken together, these studies indicate that perceived usefulness of an AWE tool and perceived writing quality may not be separate points entirely, yet they are connected empirically and theoretically.

Even though this expanding literature, most existing studies on *Grammarly* in EFL contexts tend to do not only (a) demonstrated one single overall perception score, without separating the results by specific tool feature or by essay part, or (b) investigated perception and writing-quality products as two separated-research elements, without testing the relationship between them. This leaves two gaps unaddressed. Firstly, not much is known about which particular features in *Grammarly*'s free version such as spell-checker, grammar checker, mechanics, or clarity-undergraduate EFL students see as more helpful, and which ones feel least helpful, especially in regional contexts that are sometimes under-represented, like a state university in Jambi Province. Secondly, while some studies link AWE tool perception to writing outcomes, few descriptive-quantitative studies at the Indonesian undergraduate level have examined the correlation between students' perceptions of *Grammarly* and their assessment of their own argumentative essay quality.

The present study addresses these gaps by exploring sixth-semester EFL undergraduate students' perceptions of using *Grammarly*'s free version in writing argumentative essays at a public university in Jambi Province. The inquiry examines perception at the level of both individual *Grammarly* features and argumentative-essay components and the relationship between the two variables. This research contributes both academic and practical significance. Its urgency comes from the rapid, mostly unguided implementation of freely available automated writing evaluation means such as *Grammarly* among Indonesian EFL undergraduates, a trend that has outpaced empirical evidence on which specific roles genuinely support argumentative writing and which do not, predominantly in regional, under-research

contexts for example Jambi province. By exploring more carefully at perception on the feature level and essay component, then empirically testing the relationship between the two, this study encompasses the essentially descriptive *Grammarly*-perception literature and offers a more diagnostic explanation in which the free version does and does not enhance value. Practically, the results are expected to benefit EFL writing instructors and curriculum developers, since they can use an empirical foundation for deciding which aspects of *Grammarly*-supported feedback can be beneficial and which parts still need obvious teacher guidance, for higher-order tasks for example background contextualization and argument development. Generally, the study provides context specific evidence from an Indonesian public university which could inform the responsible integration of automated writing evaluation tools into argumentative-essay instruction in similar EFL settings.

B. METHOD

This study examined undergraduate EFL students' perceptions of individual features of *Grammarly*'s free version, their perceptions of the components of their argumentative essays when supported by *Grammarly*, and the relationship between these two sets of perceptions. These objectives relate to the study's three research questions namely the first research question addresses students' perceptions of the four individual *Grammarly* features, the second question investigates their perceptions of the seven argumentative-essay components, and the third one examines the relationship between the two sets of perceptions. A quantitative, cross-sectional, descriptive-correlational survey design was adopted, consistent with common practice in EFL research on automated writing evaluation tools (Creswell & Creswell, 2017; Alnemrat et al., 2025; Hoang & Van, 2025). The descriptive component profiled students' perceptions of four *Grammarly* features (spell-checker, grammar checker, punctuation, clarity) which correspond to RQ1, seven argumentative-essay components (thesis statement, introduction, body paragraphs, counterargument and refutation, coherence, language and style, conclusion) which likn to RQ2, and the correlational component tested the association between the two sets of perceptions to answer RQ3. Data were collected at a single point in time, after respondents had completed a semester of *Grammarly*-supported argumentative essay writing.

The population comprised 145 sixth-semester students in the English Education Study Program at Universitas Jambi who had completed an Advanced Reading and Writing course. A purposive, non-probability sampling technique selected respondents who held an active *Grammarly* account, had used its free version while composing argumentative essays, and had completed relevant coursework. Slovin's formula, with a 5% margin of error, yielded a minimum required sample of 107; 107 students ultimately returned valid questionnaires (71 females, 66.4%; 36 males, 33.6%; aged 19-22, M = 20.8 years).

Data were collected using a structured, closed-ended questionnaire administered online via Google Forms, comprising 28 items rated on a four-point Likert scale (Strongly Disagree to Strongly Agree); a forced-choice format was used to minimize central-tendency bias. Ten items measured the *Grammarly* variable (X) across the four indicators above, and 18 items measured the Argumentative Essay variable (Y) across the seven indicators above, adapted from established instruments (Armanda et al., 2022; Hyland, 2019) and translated into bilingual English-Indonesian format. Content validity was established through review by three expert judges. Construct validity was examined via Pearson's product-moment correlation on a 30-respondent pilot sample, with all 28 items exceeding the r-table threshold (r-table = .191, alpha

= .05; r-count ranged from .488 to .767). Cronbach's alpha indicated high internal consistency for both the *Grammarly* (.871) and Argumentative Essay (.912) subscales.

Descriptive statistics (means, standard deviations, minimum and maximum values) addressed the perception-focused research questions, with mean scores interpreted using a four-band scale (1.00-1.74 very low; 1.75-2.49 low; 2.50-3.24 high; 3.25-4.00 very high). Prior to relational analysis, a One-Sample Kolmogorov-Smirnov test on the regression residuals confirmed the normality assumption underlying parametric testing. The relationship between the two variables was then examined using Pearson product-moment correlation and simple linear regression to determine the strength, direction, and significance ($p < .05$) of the association, with the coefficient of determination (R^2) indicating the proportion of variance in argumentative-essay perception attributable to *Grammarly* perception. All analyses were performed using SPSS.

C. FINDINGS AND DISCUSSION

Findings

Feature-Level Perceptions of Grammarly's Free-Version

To address RQ1, responses to the ten items measuring the *Grammarly* variable (Variable X) were analyzed descriptively across its four indicators: spell-checker, grammar checker, punctuation, and clarity. As shown in Table 1, the overall mean score for the *Grammarly* variable was 3.06 (SD not indicator-specific; high category), indicating that students held a generally positive perception of *Grammarly*'s free version in supporting their writing process.

Table 1. Indicator-Level Perceptions of *Grammarly*'s Free-Version Features

Indicator	Item(s)	Mean Range	Indicator Mean	Category
Spell-checker	X1–X3	2.59 – 3.14	2.94	High
Grammar checker	X4–X6	3.04 – 3.20	3.12	High
Punctuation	X7–X8	3.02 – 3.06	3.04	High
Clarity	X9–X10	3.09 – 3.28	3.19	Very High
Overall (Grammarly, Variable X)	X1–X10	2.59 – 3.28	3.06	High

At the indicator level, clarity received the highest mean score ($M = 3.19$, very high for its top-scoring item), with the single highest item-level score recorded for the statement that the clarity feature helps check the coherence of writing ($M = 3.28$). This suggests that students perceived *Grammarly*'s clarity function as particularly useful for refining the readability and coherence of their argumentative essays. In contrast, the spell-checker indicator received the lowest mean score ($M = 2.94$), with its lowest single item concerning *Grammarly*'s ability to identify and correct spelling mistakes scoring 2.59, the lowest of all 28 items in the questionnaire. Grammar checker ($M = 3.12$) and punctuation ($M = 3.04$) fell between these two extremes, both remaining within the high category. These findings indicate that while students perceive all four features of *Grammarly*'s free version favorably, their confidence in the tool is strongest for higher-order clarity support and comparatively more modest for basic spell-checking.

Components-Level Perceptions of Argumentative Essay Quality Supported by Grammarly

To answer RQ2, responses to the eighteen items measuring the Argumentative Essay variable (Variable Y) were analyzed descriptively across its seven indicators: thesis statement, introduction, body paragraphs, counterarguments and refutation, logical flow and coherence, language and style, and concluding paragraphs. The overall mean score for the Argumentative Essay variable was 3.07 (high category), indicating that students generally perceived their argumentative essays to be of high quality when supported by *Grammarly*'s free version (see Table 2).

Table 2. Indicator-Level Perceptions of Argumentative Essay Quality

Indicator	Item(s)	Mean Range	Indicator Mean	Category
Thesis statement	Y1–Y3	3.00 – 3.01	3.00	High
Introduction	Y4	2.86	2.86	High
Body paragraphs	Y5–Y8	3.01 – 3.13	3.08	High
Counterarguments and refutation	Y9–Y11	2.89 – 3.12	3.01	High
Logical flow and coherence	Y12–Y13	2.92 – 3.03	2.98	High
Language and style	Y14–Y15	3.09 – 3.20	3.15	High
Concluding paragraphs	Y16–Y18	3.10 – 3.57	3.30	Very High
Overall (Argumentative Essay, Variable Y)	Y1–Y18	2.86 – 3.57	3.07	High

Concluding paragraphs emerged as the highest-scoring indicator ($M = 3.30$, very high), with the single highest item across the entire questionnaire recorded for the statement that *Grammarly* helps produce clear and concise conclusions ($M = 3.57$). This indicates that students felt especially confident in *Grammarly*'s contribution to closing their essays effectively. Conversely, the introduction indicator received the lowest mean score ($M = 2.86$), corresponding to the item assessing whether *Grammarly* assists in providing sufficient background information to contextualize the essay — the lowest-scoring item within the Argumentative Essay variable. The remaining indicators — thesis statement ($M = 3.00$), body paragraphs ($M = 3.08$), counterarguments and refutation ($M = 3.01$), logical flow and coherence ($M = 2.98$), and language and style ($M = 3.15$) — all fell within the high category, with comparatively little spread between them. These results suggest that *Grammarly*'s free version is perceived as more supportive of surface-level and closing/summarizing aspects of argumentative writing than of the higher-order task of establishing background context in the introduction.

Relationship between Grammarly Perception and Argumentative Essay Perception

Composite scores were computed for the *Grammarly* variable (X) and the Argumentative Essay variable (Y) for all 107 respondents (see Table 3). The *Grammarly* merged scores ranged from 11 to 40 ($M = 31.18$, $SD = 5.94$), and the Argumentative Essay combined scores ranged from 27 to 70 ($M = 55.81$, $SD = 8.80$).

Table 3. Descriptive Statistics for the *Grammarly* and Argumentative Essay Composite Scores

Variable	N	Min	Max	Mean	SD
<i>Grammarly</i> (X)	107	11.00	40.00	31.18	5.94
Argumentative Essay (Y)	107	27.00	70.00	55.81	8.80

Prior to relational analysis, a One-Sample Kolmogorov–Smirnov test was conducted on the unstandardized residuals of the regression of Argumentative Essay perception on *Grammarly* perception. The residuals had a mean of .653 and a standard deviation of 6.762, with an Asymp. Sig. (2-tailed) value of .091. Because this value exceeded the .05 threshold, the residuals were considered normally distributed, satisfying the assumption required for parametric correlational analysis and supporting the use of Pearson product-moment correlation and simple linear regression to test RQ3.

Based on the reported residual variance relative to the total variance of the Argumentative Essay composite scores, the proportion of variance in Argumentative Essay perception attributable to *Grammarly* perception can be approximated at $R^2 \approx .41$, corresponding to a Pearson correlation coefficient of approximately $r \approx .64$. This indicates a moderately strong, positive association: students who perceived *Grammarly*'s free version more favorably also tended to perceive their argumentative essay writing quality more favorably.

Discussions

Perceptions of *Grammarly*'s Free-Version Features

The descriptive findings for RQ1 show that undergraduate EFL students at a state university in Jambi Province held a generally favorable but feature-differentiated perception of *Grammarly*'s free version (overall $M = 3.06$, high category), with the clarity indicator receiving the highest rating ($M = 3.19$) and the spell-checker indicator the lowest ($M = 2.94$). This pattern of uneven, feature-specific endorsement rather than uniformly high scores across all functions is consistent with, and extends, prior *Grammarly*-perception research. Dewi (2022) similarly found that EFL students perceived *Grammarly*'s evaluative function as most effective for correctness-related aspects of writing, while Armanda et al., (2022) found that students rated the tool favorably for grammar and mechanics but were more reserved about its contribution to stylistic sophistication. The present finding diverges from these two studies in one respect: although Dewi (2022) and Armanda et al., (2022) positioned surface-level correctness features as the strongest component of *Grammarly*, the sample at a state university in Jambi Province valued the clarity feature above the grammar-checker and spell-checker indicators. A believable explanation is that the free version's clarity function operates at the sentence-coherence level rather than at the level of isolated lexical or grammatical correction, and students who have already internalized basic grammar rules through prior coursework may derive moderately more perceived value from feedback that addresses how ideas are connected than from feedback that flags discrete errors.

The relatively low rating of the spell-checker indicator, and in particular of the single lowest-scoring item in the entire 28-item instrument ($M = 2.59$, concerning the tool's ability to identify and correct spelling mistakes), also merits attention. This finding is compatible with Samosir and Daulay (2023) and Komara and Putra (2024), both of whom documented feature-specific rather than globally uniform perceptions of *Grammarly* among EFL learners, and suggests that

students may regard basic spell-checking as a comparatively low-stakes function whose benefits are less salient than those of features that visibly reshape sentence-level meaning. The RQ1 findings support the view that perception of an AWE tool is not a monolithic construct; disaggregating perception by feature, as this study does, provides a more diagnostic account of where free-version *Grammarly* is perceived to add the most value than a single aggregate score would allow.

Perceptions of Argumentative Essay Quality Supported by Grammarly

The findings for RQ2 indicate that overall high perception of argumentative essay quality when supported by *Grammarly* ($M = 3.07$), with concluding paragraphs rated most favorably ($M = 3.30$, very high category) and the introduction indicator rated lowest ($M = 2.86$). This asymmetry between the strongest and weakest components maps closely onto the higher-order versus surface-level distinction that keeps coming up throughout the *Grammarly* literature. Hoang and Van (2025) reported that EFL learners perceived stronger support from *Grammarly's* feedback for surface-level correctness than for the development of argumentation, whereas Thi and Nikolov (2022) found that that EFL learners perceived stronger support from *Grammarly's* feedback for surface-level correctness, rather than for the development of argumentation, however Thi and Nikolov (2022) stated that *Grammarly* feedback complemented teacher feedback most effectively for language accuracy, with teacher input remaining indispensable for organization and argumentation.

Furthermore, the quite strong scoring of the concluding-paragraph indicator in the present study is more distinctive finding. The present sample's especially high scoring of the single item addressing clear and concise conclusion ($M = 3.57$, the highest of all 28 items) recommends that the tool's clarity and conciseness recommendations may be remarkably well-matched to the summarizing function of a conclusion, that usually includes restating already-formulated ideas instead of generating new content or evidence; but previous studies have mostly highlighted *Grammarly's* strength in correctness rather than in any specific structural element of an essay. This offers more detailed explanation compared with Inayah and Apoko's (2024) study which confirmed positive views of *Grammarly's* grammar assistance among EFL students writing analytical exposition text but noted limited support for content and structural development in the free version without identifying which structural elements were most and least affected. The RQ2 findings highlight a consistent picture across both variables since the students perceive *Grammarly's* free version as the most useful when the writing work is about refining or summarizing already-formed language. At the same time, it seems least useful when tasks require generating new content, background knowledge, or argumentative elements.

The Relationship between Grammarly Perception and Argumentative Essay Perception

The relational analysis addressing RQ3 denotes that the remains of the regression of Argumentative Essay perception on *Grammarly* perception were normally distributed (Asymp. Sig. = .091), satisfying the assumption required for parametric correlational analysis, and the two combined variables were positively and moderately strongly associated (approximate $R^2 \approx .41$ $r \approx .64$). This finding is consistent with a growing body of intervention research suggesting that perceived usefulness of an AWE tool and perceived or actual writing-quality outcomes are empirically connected instead of independent constructs. Alnemrat et al., (2025) identified that Ai-generated feedback produced writing performance gains comparable to teacher feedback on

an argumentative writing task among EFL undergraduates. Barrot (2023), furthermore, exposed that automated written corrective feedback noticeably enhanced second-language writing accuracy. The moderate-strong association observed in the Jambi sample is compatible with, though not exactly the same as, these intervention-based findings, since the present study measured perception rather than an independently scored writing outcome the relationship documented in this research should be read as an association between two sets of self-reported attitudes instead of as evidence that *Grammarly* use to improve argumentative essay quality.

This difference matters for how the RQ3 findings is positioned relative to the literature. Perception-based studies such as Dewi (2022) and Armanda et al., (2022) have found positive attitudes toward *Grammarly* without formally testing whether such attitudes co-vary with perceptions of writing quality, although intervention research conducted by Alnemrat et al., (2025) and Barrot (2023) have identified writing-quality gains without directly assessing learners' perceptions of the tool itself. Thus, by testing the association between the two constructs within a single sample, the present study addresses a gap noted in the introduction that is quite few descriptive-quantitative studies at the undergraduate level in Indonesia have investigated whether perception of an AWE-tool and perception of learners' own writing quality are significantly related. The moderate-strong positive relationship found here provides some preliminary empirical verification for that assumption, nevertheless the manuscript's methodological note also correctly reminds readers that the precise Pearson correlation, and the regressions coefficients should be re-calculated directly from the raw merged scores before submission. This is because the residual-based estimate reported here is only a preliminary approximation.

D. CONCLUSION

This study examined undergraduate EFL students' perceptions of *Grammarly*'s free version and of their own argumentative essay writing quality at a state university in Jambi Province. It, then, tested the relationship between these two kinds of perceptions. Based on survey data from 107 sixth-semester undergraduates, the descriptive results from RQ1 recommended that the students had an overall favorable yet not exactly identical view of *Grammarly*'s free version (overall $M = 3.06$, high category). In particular, the clarity related feature was perceived most positively, however the spell-checker element was perceived least positively. For RQ2, the descriptive results identified similarly favorable but still differentiated by parts, in the sense of how they viewed argumentative essay quality when supported by *Grammarly* (overall $M = 3.07$, high category). Some components in writing argumentative essay stood out for instance concluding paragraphs were scored most positively, but the introduction indicator, specifically the item about background contextualization, got the lowest scoring. Then for RQ3, the relational analysis pointed out that the residuals from the *Grammarly*-Argumentative Essay regression were normally distributed. The two combined variables were positively and moderately strongly correlated (approximate $R^2 \approx .41$, $r \approx .64$). It can be seen that learners who used *Grammarly*'s free version was more helpful, tended also perceive their argumentative essay writing quality was higher. Furthermore, the three groups of findings could be highlighted to one main that is learners generally view *Grammarly*'s free version and its support to their essays, as most beneficial in writing tasks involve polishing, making concise, or making clearer language that is already basically written.

The findings above are consistent with Hyland's (2019) discussion on argumentative writing as a genre requiring the coordinated utilization of claims, evidence, and reasoning together with surface-level accuracy, and also with Warschauer & Healey's (1998) observation that technology can support but not replace the human dimensions of writing instruction. In addition, it encompasses the feature-specific perception literature which represented by Armanda et al., (2022), Dewi (2022), Komara and Putra (2024), Samosir & Daulay (2023) by examining perception at both the feature level and essay element level, within one single sample. Some other researchers namely Hoang and Van (2025) and Thi & Nikolov (2022), similarly, test the links between the two constructs which is handled empirically in this study. This approach is in line with the intervention-based evidence found by Alnemrat et al., (2025) and Barrot (2023). The findings come with practical implications for EFL writing instruction. Because learners view *Grammarly*'s free version as differentially beneficial across features and essay components, instructors are encouraged to support its use selectively, guiding learners toward the features it support well (clarity and conciseness) while giving explicit guidance on tasks it cannot really do, for example building background context in an introduction or developing the argumentative element. The moderate-strong association between *Grammarly* perception and argumentative-essay perception also show that getting students to engage with the means does not have to diminish writing quality, as long as a complement to instructor feedback, instead of a replacement for it, especially for the higher-order parts of argumentative writing.

Several limitations should be acknowledged, firstly the study relied solely on self-reported perception data which are not independently scored writing samples. Thus, the relationship examined for RQ3 should be seen as an association between two sets of attitudes, not as proof that *Grammarly* use enhances argumentative essay quality in a causal way. Secondly, the R^2 and r values reported for RQ3 were resulted analytically from the residual and total variance of argumentative essay merged scores instead of being based on a directly computed Pearson correlation from the raw dataset that means they should be treated as preliminary estimations and ideally recalculated in SPSS using the raw merged score for Variable X and Y before submission. Thirdly, since the present study used a purposive, non-probability sampling approach and one institution cross sectional design the findings may not really generalized beyond sixth semester EFL learners at a state university in Jambi Province.

Future research could still work around these issues by pairing perception surveys with argumentative essays which are scored independently, hence it becomes clearer whether perceived usefulness can really measure writing quality improvements. It would also help to extend data collection to some institutions, to obtain stronger external validity, and to implement a longitudinal or quasi experimental design which can track changes either in *Grammarly* perception and writing quality over time. Moreover, comparative study examining *Grammarly*'s free versus premium versions, or comparing *Grammarly* with other AWE means could verify whether the characteristics specific pattern analyzed here which strong perceived support for surface level and summarizing tasks, yet weaker perceived support for background contextualization and argument development that is something uniquely tied to *Grammarly*, or more broadly a feature of AWE in general.



REFERENCES

- Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: a quantitative study. *Frontiers in Education*, 10, 1614673.
- Ananda, C. R., Mardiah, R., & Fajaryani, N. (2024). Writing Argumentative Essays: Jambi EFL Students' Challenges and Strategies. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 9(2), 339–357.
- Armanda, M. L., Nugraheni, A. F., Wulansari, A., & Imron, A. (2022). “ Grammarly” as English Writing Assistant from EFL Students' Perspective. *English Education: Journal of English Teaching and Research*, 7(2), 128–137.
- Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584–607.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Dewi, U. (2022). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. *Lingua Cultura*, 16(2), 155–161.
- Ghanbari, N., & Salari, M. (2022). Problematizing argumentative writing in an Iranian EFL undergraduate context. *Frontiers in Psychology*, 13, 862400.
- Hassanzadeh, M., & Fotoohnejad, S. (2021). Implementing an automated feedback program for a Foreign Language writing course: A learner-centric study: Implementing an AWE tool in a L2 class. *Journal of Computer Assisted Learning*, 37(5), 1494–1507.
- Hoang, G. T. L., & Van, K. T. G. (2025). Grammarly Feedback on EFL Learners' Writing: Feedback Precision and Student Perceptions. *EIKI Journal of Effective Teaching Methods*, 3(2).
- Hyland, K. (2019). *Second language writing*. Cambridge university press.
- Inayah, T. M., & Apoko, T. W. (2024). Exploring students' perspectives on the use of Grammarly in writing analytical exposition text. *JLE: Journal of Literate of English Education Study Program*, 5(1), 73–83.
- Komara, C., & Putra, A. R. (2024). STUDENTS' PERCEPTION OF USING GRAMMARLY WEBSITE TO SUPPORT THEIR ENGLISH WRITING SKILL. *Klasikal: Journal of Education, Language Teaching and Science*, 6(2), 265–272.
- Ozfidan, B., & Mitchell, C. (2020). Detected difficulties in argumentative writing. *Journal of Ethnic and Cultural Studies*, 7(2), 15–29.
- Pasaribu, A. N., Pasaribu, T. K., Siahaan, R. B., & Sitompul, D. S. (2024). Challenges for Indonesian EFL students in English academic writing. *English Review: Journal of English Education*, 12(2), 527–536.
- Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674.
- Samosir, S. A., & Daulay, S. H. (2023). EFL students' perception of the use “Grammarly application” as Automated Writing Evaluation (AWE). *Conference on English Language Teaching*, 928–940.
- Saricaoglu, A. (2019). The impact of automated feedback on L2 learners' written causal explanations. *ReCALL*, 31(2), 189–203.
- Thi, N. K., & Nikolov, M. (2022). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher*, 31(6), 767–779.
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview.



Language Teaching, 31(2), 57–71.

- Xie, Y., & Lv, X. (2022). Effects of content support and planning instruction on discourse connection in EFL argumentative writing. *Frontiers in Psychology*, 13, 912311.
- Xu, J., & Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *The Asia-Pacific Education Researcher*, 31(4), 357–367.
- Yu, R., & Yang, L. (2021). ESL/EFL learners' responses to teacher written feedback: Reviewing a recent decade of empirical studies. *Frontiers in Psychology*, 12, 735101.