

The Role of Grammarly in Saudi EFL Writing: Students' Perceptions of Its Effectiveness, Efficiency, and Overreliance

**Ahmed Mohammed Ahmed Abdullah, Saleem Mohd Nasim
and
Minahi Abdullah Alqahtani**

Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

The rapid adoption of AI writing assistants such as Grammarly in education has influenced how English language learners approach writing tasks. By providing automated feedback on linguistic errors, including mechanics (spelling, punctuation, and capitalization), grammar, and style, these tools may help learners improve writing accuracy, reduce cognitive load, and write more efficiently. However, excessive reliance on such tools is a cause for concern. This study investigated Saudi male preparatory-year EFL students' perceptions of Grammarly at Prince Sattam bin Abdulaziz University (n = 61). A validated questionnaire was used in a cross-sectional design to collect data on perceived effectiveness in correcting writing mechanics (spelling, capitalization, and punctuation), perceived efficiency, limitations, and overreliance. The findings indicated that most students perceived Grammarly as useful in correcting writing mechanics, particularly spelling (85.2%), capitalization (78.7%), and punctuation (68.9%). In addition, 68.8% of students agreed that Grammarly saved time and effort. Responses regarding overreliance were mixed: 37.7% indicated high overreliance on Grammarly, whereas 62.3% reported low/no overreliance. Spearman's rho showed no significant association between perceived usefulness and overreliance ($\rho = .105$, $p = .419$). These findings suggest that Grammarly can be integrated into EFL writing instruction as a scaffold to support writing mechanics, provided that its use is guided in ways that also promote critical engagement and learner autonomy.

Keywords: artificial intelligence (AI) writing assistants; grammarly; Saudi EFL writing; writing mechanics; overreliance; learner autonomy; EFL writing instruction

The recent rise of artificial intelligence (AI) has had a significant impact on various aspects of life, including foreign language instruction. AI-powered writing assistants, such as Grammarly, have transformed traditional writing practices. For example, Grammarly can help users generate ideas, develop them into coherent content, and identify errors by providing instant feedback on critical aspects of writing, including spelling, capitalization, and punctuation. Studies support its efficacy. Yawiloeng (2025) and Mohd Sham et al. (2025) demonstrated that Grammarly improved grammatical accuracy and writing fluency. Similarly, Estrella (2025) and Ghuftron and Rosyida (2018) highlighted

Grammarly's role in fostering greater independence and efficiency in writing. Collectively, these findings suggest that Grammarly has played an important role in shaping how English as a Foreign Language (EFL) learners develop writing skills.

In the context of Saudi Arabian education, which is undergoing rapid modernization, tools like Grammarly hold significant promise and occupy a unique niche. Preparatory-year university students, who are predominantly Gen Z (aged 17–22) and highly tech-savvy, increasingly rely on AI tools to enhance their writing skills (Ozfidan et al., 2024). Grammarly, in particular, has been recognized for helping Saudi students and other non-native speakers address limitations in language proficiency (Khan et al., 2024). However, the successful implementation of educational technology depends not only on its technical efficiency but also on how users perceive it (Nasim et al., 2024). Understanding whether students view Grammarly as a valuable scaffold for learning or merely as a convenient crutch is critical to evaluating its overall impact on the educational landscape.

Studies indicate that English as a Foreign Language (EFL) students face notable challenges in writing, especially with mechanics such as spelling, punctuation, and capitalization (Nasim & Mujeeba, 2024). Additionally, presenting and understanding ideas clearly poses significant difficulties for many learners. Recent Saudi-context research also suggests that students' perceptions of writing problems do not always fully align with their actual written performance (Ishtiaq et al., 2025). While tools like Grammarly have become popular for addressing such issues, their focus is primarily on achieving mechanical accuracy rather than improving argumentation or coherence (Muhammad, 2024; Resiana et al., 2024). Existing literature consistently highlights the technical effectiveness of Grammarly in correcting errors (Fitria & Sabarun, 2022; Fitria, 2023; Hoang & Van, 2025; Huang et al., 2020). However, there has been little focus on understanding students' perceptions of Grammarly's effectiveness, efficiency, limitations, and the risks of overreliance, particularly in the context of Saudi Arabian preparatory-year EFL programs.

This study addresses that gap by examining Saudi preparatory-year EFL students' perceptions of Grammarly in relation to four connected areas: its effectiveness in correcting writing mechanics, its efficiency in saving time and effort, its perceived limitations in fully correcting errors, and students' possible overreliance on the tool. It also investigates whether perceived usefulness is significantly related to perceived overreliance. By adopting this user-centered and context-specific approach, the study aims to provide a more precise understanding of Grammarly's role in Saudi EFL writing and to offer evidence that can inform pedagogy, instructional design, and responsible integration of AI-assisted writing support.

The following research questions guide the present study:

- RQ1.** To what extent do preparatory-year Saudi EFL students perceive Grammarly as effective in correcting writing mechanics (spelling, capitalization, and punctuation)?
- RQ2.** What are students' perceptions of Grammarly's efficiency in saving time and effort and its limitations in fully correcting writing errors?
- RQ3.** To what extent do students perceive their overreliance on Grammarly when correcting writing errors in writing mechanics?
- RQ4.** Is there a statistically significant relationship between students' perceptions of Grammarly's usefulness (mechanics correction and efficiency) and their perceived overreliance on Grammarly?

By addressing these research questions, this study aims to provide a more detailed perspective on the role of Grammarly in the Saudi EFL context and to inform pedagogical practice and policy.

Literature Review

The Efficacy of Grammarly in Error Correction

EFL education is increasingly influenced by the accelerated adoption of AI. The technical effectiveness of AI writing assistants such as Grammarly in detecting and correcting major and minor EFL writing errors has been reported in previous studies. Hoang and Van (2025) and Fitria (2023) assert that Grammarly accurately identifies capitalization, spelling, punctuation, and grammar errors. In other words, students who used Grammarly not only made fewer errors in writing mechanics but also achieved better writing outcomes, showing improvements in general writing proficiency and suggesting that iterative interaction with AI feedback could lead to broader linguistic growth (Estrella, 2025; Resiana et al., 2024; Martinez-Carrasco & Chabert, 2023).

Similarly, Wahyu and Zur (2023) Resiana et al., (2024) demonstrated that Grammarly contributed to improving various aspects of writing, such as grammar, spelling, punctuation, and overall readability. In line with this result, Dizon and Gayed (2021) emphasized the idea that Grammarly helps L2 learners write grammatically correct texts that facilitate effective communication in a second language (L2). Additionally, Dewi (2023) found that Grammarly offered clear and easy-to-understand feedback and saved time for learners when assessing grammar, verb use, word order, passive phrases, and impersonal constructions. This suggests that Grammarly may be a useful tool in developing basic writing skills among preparatory-year students who are learning the fundamentals of academic writing.

Learner Perceptions: Affordances, Challenges, and Nuanced Realities

The literature presents not only quantitative indicators of effectiveness but also a complex landscape of learner perceptions. Learners reported the affective advantages of AI-based feedback, such as less frustration and more confidence, because the immediacy of AI-based feedback eases the cognitive burden of the writing process (Mekheimer, 2025; Bahing et al., 2025; Lai, 2025; Resiana et al., 2024). Furthermore, perceived time efficiency is another consistently cited advantage, allowing learners to allocate more cognitive resources to higher-order concerns like organization and argumentation (Al Mahmud, 2023; Altamimi, 2025). This suggests that learners may value Grammarly not only for correction itself but also for the way it supports a more manageable and less demanding writing process.

However, recent studies also indicate that while Grammarly is relatively effective in correcting simple mechanics, it may be less effective in addressing more complex or contextual writing errors, particularly in areas that require context-sensitive judgment (Abu Guba et al., 2024). Additionally, concerns about overdependence on automated feedback have also emerged. Researchers caution that this may result in disengagement from the writing process and a gradual sense of deskilling, where learners unconsciously accept corrections without critically analyzing the underlying linguistic principles, which can impede the development of their own autonomous proofreading abilities (Hoang & Van, 2025; Vieriu & Petrea, 2025). This dichotomy—between recognized benefits and risks of reliance—reinforces the notion that learners' perceptions are not a unitary construct but a balance of recognized affordances and acknowledged risks.

Contextualizing the Gap: The Saudi EFL Learner

Although the existing body of research on AI tools like Grammarly offers a valuable framework, its direct relevance and applicability to the Saudi preparatory-year context remain limited. Recent Saudi-context studies also show that punctuation, capitalization, and grammar remain frequent writing problems, and that students' perceptions do not always fully align with their actual writing practices (Nasim & Mujeeba, 2024; Ishtiaq et al., 2025). The adoption of AI and its various forms in the Saudi educational context is shaped by Saudi Vision 2030, yet only a few studies have investigated

the advantages, challenges, and time efficiency of Grammarly specifically within this context and population. Existing studies in other contexts have established the general dimensions of perception, such as advantages and disadvantages, but they have not systematically investigated the interrelationships among efficacy, efficiency, and overuse, nor have they examined students' perceptions in the context of Saudi foundation programs. This research addresses this gap by using a validated instrument to examine these perceptions, thereby providing a more context-specific analysis of Grammarly's role among Saudi students. By synthesizing these themes, this study highlights the need for a perception-driven account of Grammarly's integration, which is important for effective, context-sensitive pedagogical interventions.

Method

Research Design

The present study employed a quantitative survey design to examine preparatory-year Saudi EFL students' perceptions of Grammarly's effectiveness, efficiency, limitations, and possible overreliance. The research followed a descriptive-correlational approach to examine both individual dimensions of Grammarly use and the relationships among these dimensions. The study focused on how Saudi EFL students perceive Grammarly as a tool for correcting mechanical errors in writing. A quantitative methodology was considered appropriate because it enabled the systematic analysis of responses related to the use of AI writing assistants.

The methodological framework combined descriptive and inferential statistical methods to answer the research questions while maintaining alignment with the study's overall aim.

Participants and Sampling

The research was conducted using a census of all accessible preparatory-year male EFL students at Prince Sattam bin Abdulaziz University during the semester of the academic year 2025-26, resulting in a final sample of 61 students. The participants were Saudi EFL students aged 17 to 20 years whose first language was Arabic. All participants were enrolled in the first semester of a Level 1 English language writing course at the time of the study, which ensured consistent foundational exposure to academic writing instruction and Grammarly use.

Participants were included based on the following criteria: enrollment in the preparatory-year English program, use of Grammarly during the academic semester for writing tasks, voluntary participation in the survey, and completion of all questionnaire items. The institution was a public university in Saudi Arabia that enrolled preparatory-year EFL learners with A1-A2 proficiency levels based on the CEFR. Participants were also familiar with digital technologies through prior exposure to educational technology.

Instrument Development and Validation

The data collection instrument was a structured, closed-ended questionnaire developed for the study. It included three items measuring perceived effectiveness in correcting spelling, capitalization, and punctuation (items 1-3), one item measuring perceived efficiency in saving time and effort (item 4), one item measuring perceived limitations in overall error correction (item 5), and one item measuring perceived overreliance on Grammarly (item 6). The questionnaire used a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

To ensure clarity and relevance, the instrument was reviewed by three EFL experts. Face validity was further examined through pilot testing with 10 students, after which minor revisions were

made based on their feedback. Reliability testing showed satisfactory psychometric properties, with Cronbach's alpha of 0.84 for the effectiveness subscale, which exceeded the acceptable threshold of 0.70. Item-total correlations ranged from 0.54 to 0.72, indicating adequate internal consistency.

Data Collection Procedure

Data were collected after approval had been obtained from the relevant departmental authorities. All participants were informed about the purpose of the study and were assured that their responses would remain confidential through anonymized data collection and storage procedures. Participation was voluntary, and no penalties were associated with non-participation. The questionnaire was created using Google Forms and distributed to participants via WhatsApp. The researchers kept the survey open for two weeks. Quality-control procedures included data coding, data cleaning, and verification of data accuracy.

Data Analysis

Data were analyzed using IBM SPSS Statistics 27. The analysis began with screening for missing data and assessment of normality using the Shapiro-Wilk and Kolmogorov-Smirnov tests, along with skewness and kurtosis values. Descriptive and inferential statistical techniques were then used to address the research questions. The data analysis framework is presented in Table 1.

Table 1

Statistical Analysis Plan by Research Question

Research Question	Focus	Primary Analytical Methods	Rationale / Key Metrics
RQ1: Effectiveness	Perceived effectiveness across writing mechanics	<ul style="list-style-type: none"> • Descriptive Statistics • Binomial Tests • Friedman Test • Post-hoc Wilcoxon signed-rank tests 	Compare effectiveness percentages with the reference value (50%), test differences across spelling, capitalization, and punctuation, and assess internal consistency of the effectiveness scale.
RQ2: Efficiency & Limitations	Perceptions of time-saving efficiency and perceived limitations	<ul style="list-style-type: none"> • Reliability Analysis (Cronbach's α) • One-sample Wilcoxon signed-rank test • Cross-tabulation with Chi-square & Phi • Descriptive Comparisons 	Test whether efficiency and limitation scores differ from the neutral point and examine the association between perceived efficiency and perceived limitations.
RQ3: Overreliance	Level and distribution of perceived dependency	<ul style="list-style-type: none"> • Distribution Analysis (Skewness, Kurtosis, Percentiles) • One-sample Wilcoxon signed-rank test • Binomial tests • Categorical frequency analysis 	Examine whether overreliance scores differ from the neutral point, test the proportion reporting high overreliance, and describe the distribution of responses.
RQ4: Relationship	Relationship between perceived usefulness and perceived overreliance	<ul style="list-style-type: none"> • Spearman's Rank Correlation (ρ) • Mann-Whitney U Test • Cross-tabulation with Chi-square 	Test the relationship between usefulness and overreliance and compare usefulness scores across overreliance groups.
Overall Statistical Model	Guiding statistical decisions	<ul style="list-style-type: none"> • Tests: Non-parametric tests (Wilcoxon, Friedman, Mann-Whitney U, Spearman's ρ) • Effect Sizes: Kendall's W, Phi, Spearman's ρ • Multiple Comparison Control: Bonferroni Correction 	Appropriate for ordinal Likert-scale data and non-normal distributions; provides robust estimates and controls Type I error.

Non-parametric tests were used for all inferential analyses because the data were ordinal and the distributions deviated from normality. Effect sizes were reported for all tests, and Bonferroni-adjusted significance levels were applied where multiple comparisons were conducted.

Results

The findings are presented according to the four research questions and align with the study's focus on perceived effectiveness, efficiency, limitations, overreliance, and the relationship between usefulness and overreliance.

RQ1: Perceived Effectiveness in Correcting Writing Mechanics

Reliability and Internal Consistency

The three effectiveness items demonstrated good internal consistency (Cronbach's $\alpha = .84$), supporting their use as a coherent scale for measuring perceived effectiveness. Inter-item correlations ranged from .538 to .724, and all corrected item-total correlations exceeded .64, indicating strong relationships among the three items.

Table 2

Descriptive Statistics and Distribution of Perceptions and Binomial Test Results for Effectiveness Perceptions in Correcting Writing Mechanics (n=61)

Mechanics Domain	M	Mdn	SD	SD % (1)	D % (2)	N % (3)	A % (4)	SA % (5)	Not Effective % (1-3)	Effective % (4-5)	Binomial Test p-value
Spelling	4.13	4	0.97	3.3	4.9	6.6	45.9	39.3	14.8	85.2	p < .001
Capitalization	4.08	4	0.99	3.3	3.3	4.8	39.3	39.3	21.3	78.7	p < .001
Punctuation	3.87	4	1.09	4.9	4.9	21.3	36.1	32.8	31.1	68.9	p = .004

Note: Test proportion = 0.50. All tests were two-tailed.

Table 3

Friedman Test and Post-hoc Pairwise Comparisons

Analysis	Test Statistic	df	p	Effect Size
Friedman Test	$\chi^2 = 6.44$	2	0.04	Kendall's W = 0.053
Pairwise Comparisons	Z	Asymp. Sig.	Adj. Sig.*	
Punctuation vs. Capitalization	-1.583	0.113	0.773	
Punctuation vs. Spelling	-2.514	0.012	0.339	
Capitalization vs. Spelling	-0.544	0.586	1	

Note: Significant at $p < .05$ before Bonferroni correction; not significant after adjustment ($p > .0167$)

One-sample binomial tests (see Table 2) showed that a significantly greater proportion of students perceived Grammarly as effective across all three areas of writing mechanics: spelling (85.2% vs. 14.8%, $p < .001$), capitalization (78.7% vs. 21.3%, $p < .001$), and punctuation (68.9% vs. 31.1%, $p = .004$). These results indicate generally positive perceptions of Grammarly's effectiveness in correcting writing mechanics among preparatory-year Saudi EFL students.

A Friedman test (see Table 3) indicated a statistically significant overall difference in perceived effectiveness across spelling, capitalization, and punctuation, $\chi^2 (2) = 6.44$, $p = .040$. However, the effect size was very small (Kendall's $W = 0.053$). Post-hoc pairwise comparisons with Bonferroni adjustment showed that none of the individual comparisons remained statistically significant after correction.

The identical medians across the three items (Mdn = 4.00) and the negative skewness values for all items suggest that students' perceptions were generally favorable across spelling, capitalization, and punctuation, although punctuation received slightly lower ratings than the other two areas. Overall, these findings suggest that students perceived Grammarly as most effective for spelling, followed closely by capitalization, while still viewing it positively for punctuation.

RQ2: Perceptions of Grammarly's Efficiency and Limitations

This section examines students' perceptions of Grammarly's efficiency in saving time and effort and its limitations in fully correcting writing errors. As shown in Table 4, students reported positive perceptions of Grammarly's efficiency, whereas their views on its limitations were more mixed.

Table 4

Descriptive Statistics for Efficiency and Limitations Perceptions and One-Sample Wilcoxon Signed-Rank Tests Against Neutral Point (3) (n=61)

Variable	M	Mdn	SD	Distribution of Responses (%)	Test Statistic	Standardized Test Statistic	p-value	Decision
Efficiency (Item 4)	4.03	4	0.95	1.6% (SD), 1.6% (D), 27.9% (N), 29.5% (A), 39.3% (SA)	948	5.463	< .001	Reject null hypothesis
Limitations (Item 5)	3.08	3	1.13	11.5% (SD), 14.8% (D), 37.7% (N), 26.2% (A), 9.8% (SA)	400	0.446	0.656	Retain null hypothesis

Table 5

Association Between Efficiency and Perceived Limitations

Analysis	χ^2	p	Phi	Interpretation
Efficiency \times Limitations	0.241	0.624	0.063	No significant association

The one-sample Wilcoxon signed-rank test (see Table 4) indicated that students' ratings of Grammarly's efficiency were significantly above the neutral point (Test Statistic = 948.000, Standardized Test Statistic = 5.463, $p < .001$). The mean rating was 4.03, and 68.8% of students agreed or strongly agreed that Grammarly saved time and effort. This finding suggests that students generally perceived Grammarly as efficient in supporting writing tasks.

In contrast, perceptions of Grammarly's limitations did not differ significantly from the neutral point (Test Statistic = 400.000, Standardized Test Statistic = 0.446, $p = .656$). The mean rating was 3.08, with 37.7% of students selecting the neutral option, 26.3% disagreeing or strongly disagreeing, and 36.0% agreeing or strongly agreeing. This pattern indicates mixed views regarding Grammarly's ability to fully correct writing errors and suggests that no clear consensus emerged on this issue.

Relationship Between Efficiency and Perceived Limitations

Cross-tabulation analysis of binarized efficiency and limitations variables revealed no significant association between perceptions of Grammarly's efficiency and perceptions of its limitations, $\chi^2(1) = 0.241$, $p = .624$, $\Phi = .063$. The small effect size indicates that students who perceived Grammarly as highly efficient were not significantly more or less likely to perceive limitations (see Table 5).

The crosstabulation further showed that among students who perceived high efficiency (68.9% of the sample), 38.1% also agreed that Grammarly had limitations, whereas among those who perceived lower efficiency (31.1% of the sample), 31.6% reported such limitations. This pattern suggests that perceptions of limitations were not significantly associated with perceptions of efficiency.

Overall, these findings suggest that while Saudi EFL students valued Grammarly's time-saving benefits, they held mixed views about its limitations. In other words, perceiving Grammarly as efficient did not necessarily correspond to stronger or weaker perceptions of its limitations.

RQ3: Perceived Overreliance on Grammarly

This section examined the extent to which preparatory-year Saudi EFL students perceived themselves as overreliant on Grammarly when correcting mechanical writing errors. Overall, the findings indicate mixed perceptions, with some evidence of overreliance but no clear majority endorsing high dependence on the tool.

Table 6
Descriptive Statistics for Perceived Overreliance

Statistic	Value	Interpretation
Mean	3.31	Slightly above neutral
Median	3.00	Neutral point
Standard Deviation	1.07	Moderate variability
Skewness	0.092	Approximately symmetric distribution
Response Distribution	3.3% SD, 18.0% D, 41.0% N, 19.7% A, 18.0% SA	Mixed responses

As shown in Table 6, descriptive statistics indicated a mean overreliance score of 3.31, a median of 3.00, and a standard deviation of 1.07. The response distribution further showed that the largest category was neutral (41.0%), followed by agree (19.7%) and strongly agree (18.0%), whereas 18.0% disagreed and 3.3% strongly disagreed. The skewness value (0.092) suggests that the distribution was approximately symmetric, indicating that responses were mixed rather than concentrated at the high end of the scale.

Table 7
Inferential Results for Perceived Overreliance

Analysis	Result	<i>p</i>	Interpretation
One-sample Wilcoxon signed-rank test	$T = 474.00, Z = 2.30$.021	Significant difference from neutral point
Binomial test (High Overreliance vs. Low/No Overreliance)	37.7% high overreliance vs. 62.3% low/no overreliance	.072	No significant majority reporting high overreliance

Note. Low/No Overreliance = responses 1 to 3; High Overreliance = responses 4 to 5.

As presented in Table 7, a one-sample Wilcoxon signed-rank test indicated a statistically significant difference from the neutral point, $T = 474.00, Z = 2.30, p = .021$. However, this result should be interpreted alongside the descriptive pattern in Table 6, in which the median remained exactly at the neutral point and neutral responses constituted the largest single category.

To provide additional clarity, responses were re-coded into two categories: Low/No Overreliance (responses 1 to 3) and High Overreliance (responses 4 to 5). As shown in Table 7, 62.3% of students fell into the Low/No Overreliance category, whereas 37.7% fell into the High Overreliance category. The binomial test indicated that the proportion reporting high overreliance did not differ significantly from the reference proportion of .50, $p = .072$. Taken together, the results in Tables 6 and 7 suggest that students' perceptions of overreliance on Grammarly were mixed and that the overall pattern does not indicate a clear majority endorsing high dependence.

RQ4: Relationship Between Perceived Usefulness and Overreliance

This section examined whether students' perceived usefulness of Grammarly was significantly related to their perceived overreliance on the tool. Overall, the findings showed no statistically significant relationship between these two variables.

Table 8

Descriptive Statistics for Usefulness and Overreliance Variables and Correlation Analysis Between Them (n=61)

Variable	M	Mdn	SD	IQR	95% CI	Correlation Coefficient	p	Interpretation
Perceived usefulness composite (mechanics + efficiency)	4.03	4	0.83	1	[3.82, 4.24]	Spearman's rho = .105	0.419	Small, non-significant relationship
Overreliance	3.31	3	1.07	1	[3.04, 3.59]	—	—	—

As shown in Table 8, students reported generally high perceptions of Grammarly's usefulness ($M = 4.03$, $SD = 0.83$), whereas perceptions of overreliance were more moderate ($M = 3.31$, $SD = 1.07$). Spearman's rank-order correlation revealed a small, non-significant positive relationship between perceived usefulness and overreliance, $\rho = .105$, $p = .419$. This suggests that students who perceived Grammarly as more useful were only slightly more likely to report higher overreliance, but this tendency was not statistically significant.

Table 9

Group Comparison: Usefulness by Overreliance Category

Overreliance Group	n	Mean Usefulness	Median Usefulness	Standard Deviation
Low/No Overreliance	38	3.95	4	0.87
High Overreliance	23	4.15	4.25	0.77
Total	61	4.03	4	0.83

Note. Mann-Whitney $U = 366.00$, $Z = -1.07$, $p = .285$, $r = .14$. The effect size was calculated as $r = Z / \sqrt{N}$.

As presented in Table 9, students in the High Overreliance group reported slightly higher usefulness scores ($M = 4.15$, $SD = 0.77$) than those in the Low/No Overreliance group ($M = 3.95$, $SD = 0.87$). However, a Mann-Whitney U test showed that this difference was not statistically significant, $U = 366.00$, $Z = -1.07$, $p = .285$, $r = .14$.

A supplementary cross-tabulation of categorized usefulness levels by overreliance group also showed no significant association, $\chi^2(2) = 2.052$, $p = .358$, $\Phi = .183$. Although students in the High Overreliance group were somewhat more represented in the high-usefulness category (52.2% vs. 42.1%), and students in the Low/No Overreliance group were more represented in the low-usefulness category (28.9% vs. 13.0%), these differences did not reach statistical significance.

Taken together, the results in Tables 8 and 9 suggest that, in this sample, students' perceived usefulness of Grammarly was not significantly associated with their perceived overreliance.

Discussion

This study examined Saudi preparatory-year EFL students' perceptions of Grammarly's effectiveness in correcting writing mechanics, its efficiency, its limitations, possible overreliance, and the relationship between perceived usefulness and overreliance. Overall, the findings suggest that students viewed Grammarly positively as a support tool for mechanical accuracy and writing efficiency, while also reporting mixed views about its limitations and the extent to which they relied on it. These results position Grammarly as a useful aid in EFL writing, but not as a substitute for learner judgment or teacher guidance.

Effectiveness of Grammarly in Correcting Writing Mechanics

The findings for RQ1 indicate that students generally perceived Grammarly as effective in correcting writing mechanics, especially spelling and capitalization. This pattern is consistent with previous research showing that Grammarly performs well in identifying and correcting surface-level errors such as spelling, capitalization, punctuation, and grammar (Fitria, 2023; Hoang & Van, 2025; Huang et al., 2020). It is also in line with studies suggesting that Grammarly can contribute to broader improvement in writing quality and accuracy through iterative engagement with automated feedback (Estrella, 2025; Resiana et al., 2024; Martinez-Carrasco & Chabert, 2023). The strong endorsement of spelling and capitalization in the present study suggests that students saw Grammarly as particularly useful for more rule-governed aspects of writing, which are often challenging for EFL learners. This is especially relevant in the Saudi EFL context, where writing mechanics remain a persistent source of difficulty. Nasim and Mujeeba (2024) report that Arab EFL learners commonly struggle with spelling, punctuation, and capitalization, while Ishtiaq et al., (2025) similarly found that punctuation, capitalization, and grammar were among the most common difficulties in Saudi learners' actual written performance.

Punctuation was also viewed positively, but it received slightly lower ratings than spelling and capitalization. This difference should be interpreted cautiously because the overall Friedman effect was very small and the post-hoc comparisons were not significant after Bonferroni adjustment. Still, the direction of the findings suggests that students perceived Grammarly as somewhat less dependable for punctuation than for the other two mechanics domains. This interpretation is consistent with Abu Guba et al., (2024) who argued that Grammarly may be less effective in addressing more context-dependent writing issues than simpler mechanical errors. It also resonates with Saudi and Arab EFL studies showing that punctuation remains a particularly problematic area for learners. Nasim and Mujeeba (2024) found that students and instructors did not fully agree on which mechanics errors were most serious, with students emphasizing spelling and instructors emphasizing punctuation, while Ishtiaq et al., (2025) found that punctuation errors were among the most prominent problems in Saudi learners' writing. Taken together, these converging findings support the conclusion that Grammarly is especially useful for rule-based correction, but somewhat less dependable where contextual judgment is more heavily involved.

Overall, the RQ1 findings support the view that Grammarly is most useful as a tool for mechanical support rather than as a replacement for human judgment in broader aspects of writing such as coherence, argumentation, and style (Muhammad, 2024; Resiana et al., 2024).

Efficiency and Perceived Limitations

The findings for RQ2 show that students clearly perceived Grammarly as efficient in saving time and effort. This result is consistent with previous studies that identify time efficiency as one of the most frequently reported benefits of AI-based writing feedback (Al Mahmud, 2023; Altamimi, 2025; Yang, 2018). In practical terms, this suggests that students viewed Grammarly as useful for reducing the burden of lower-level editing tasks, thereby allowing greater attention to idea development and text organization. This interpretation is also compatible with earlier studies showing that learners often value Grammarly for its convenience, clarity of feedback, and support for a more manageable writing process (Dizon & Gayed, 2021; Dewi, 2023).

At the same time, students' views about Grammarly's limitations were mixed. The non-significant Wilcoxon result indicates that there was no strong consensus about whether Grammarly fully corrects writing errors. This is an important result because it suggests that students were not simply idealizing the tool. Instead, they appeared to value its usefulness while remaining uncertain about its completeness. The non-significant association between efficiency and perceived limitations

further strengthens this interpretation. Students who viewed Grammarly as efficient were not significantly more or less likely to perceive limitations. In other words, perceiving Grammarly as helpful for saving time did not prevent students from also recognizing that it may not fully resolve all writing problems.

This interpretation fits well with the broader Saudi and Arab EFL writing literature. Nasim and Mujeeba (2024) emphasize that understanding learners' and instructors' perceptions of mechanics is important for targeted instructional design, while Ishtiaq et al., (2025) report that Saudi learners' views about their writing problems do not always align neatly with their actual writing practices. Together, these studies support the idea that learners can hold layered and even partially conflicting views about writing support tools. This point is also consistent with Nasim et al., (2024), who argue that pedagogical perspectives influence learning, development, and implementation. The present findings therefore suggest that Grammarly's efficiency and its limitations should be treated as coexisting perceptions rather than mutually exclusive positions.

Perceived Overreliance on Grammarly

The findings for RQ3 indicate mixed perceptions of overreliance. Although the one-sample Wilcoxon test suggested a statistically significant difference from the neutral point, the descriptive pattern remained more cautious: the median was exactly neutral, the neutral category was the largest single response, and the binomial test showed no significant majority reporting high overreliance. Taken together, these results suggest that some students perceived themselves as relying heavily on Grammarly, but that high overreliance was not the dominant pattern in the sample.

This more cautious interpretation still fits the broader literature warning that AI tools can create dependence if used uncritically (Hoang & Van, 2025; Vieriu & Petrea, 2025), but it also suggests that such dependence should not be assumed to characterize most learners in this context. In the present study, Grammarly appears to function as a scaffold for many students, while still presenting a possible reliance risk for a smaller group. This point becomes more meaningful when read alongside the local writing context reflected in Saudi and Arab studies. Both Nasim and Mujeeba (2024) and Ishtiaq et al., (2025) show that mechanics remain a real source of difficulty for learners, especially in punctuation, capitalization, and spelling. In a context where learners continue to struggle with these areas, some degree of reliance on AI-assisted correction is unsurprising. The important point, however, is that the current data do not support the claim that most students had become highly dependent on Grammarly.

The present interpretation also supports the view that overdependence should be treated as a possible risk rather than as an outcome that can be presumed. Studies such as Hoang and Van (2025) and Vieriu and Petrea (2025) caution against uncritical acceptance of automated feedback because it may reduce active engagement with the writing process and weaken proofreading autonomy. The current findings support that concern in a limited way, but they do not justify the stronger claim that overreliance was widespread in this sample.

Relationship Between Perceived Usefulness and Overreliance

The findings for RQ4 showed no statistically significant relationship between perceived usefulness and overreliance. The Spearman correlation was small and non-significant, and the Mann-Whitney group comparison and supplementary cross-tabulation also failed to show significant differences. This means that students who viewed Grammarly as more useful were not significantly more likely to report higher overreliance. This is an important finding because it suggests that positive perceptions of usefulness do not automatically translate into problematic dependence. In this sample, students rated Grammarly highly for usefulness, yet this did not correspond to a significant increase in perceived overreliance.

The RQ4 findings support the view that usefulness and overdependence should not be treated as identical constructs. The safest conclusion is that the relationship appears weak and non-significant in this dataset rather than entirely absent. This interpretation is also consistent with research suggesting that learners' reliance on AI tools is shaped by multiple factors, including self-regulation, metacognitive awareness, digital literacy, and instructional context, rather than by perceived usefulness alone (Rafida et al., 2024; Vieriu & Petrea, 2025; Lai, 2025). In this sense, the present results contribute to a more differentiated understanding of Grammarly use in EFL writing: students may recognize the tool's utility without necessarily becoming overly dependent on it.

Theoretical and Practical Implications

Theoretically, the findings suggest that perceived usefulness and perceived overreliance should be treated as related but distinct constructs in studies of educational technology. In this study, students rated Grammarly highly for usefulness, yet this did not correspond to a significant increase in perceived overreliance. This implies that models of AI tool use in educational settings may benefit from incorporating learner agency, critical digital literacy, and instructional scaffolding alongside more traditional acceptance variables. In this respect, the findings are compatible with broader arguments in the AI-feedback literature that learner experience, confidence, and resilience mediate how automated support affects writing development (Lai, 2025; Mekheimer, 2025).

Practically, the findings support the guided integration of Grammarly in EFL writing instruction. Grammarly can be positioned as a support tool for final-stage editing of mechanical errors rather than as a replacement for drafting and idea development. Instructors can also use the tool's limitations, especially in areas such as punctuation and contextual correction, as opportunities to encourage students to evaluate suggestions rather than accept them automatically. This is especially important because metacognitive awareness supports language learning development and helps learners regulate, monitor, and reflect on their own learning processes (Mohd Nasim, 2022). In addition, because students reported mixed perceptions of overreliance, teachers should monitor use patterns and encourage balanced engagement with AI-based writing support rather than assuming either complete dependence or complete autonomy.

Nasim and Mujeeba (2024) emphasize that understanding perceptions of mechanics can help instructors design more targeted support, while Ishtiaq et al. (2025) highlight the role of practice, feedback, and reading habits in improving writing performance. Nasim et al., (2024) further underline the importance of pedagogical perspectives in shaping implementation and learning outcomes. Together, these studies support a pedagogy in which Grammarly is integrated as one tool within a broader writing-development process rather than treated as the main driver of improvement. This is also compatible with Saudi preparatory-year evidence showing that instructional practices, learning environments, and materials are more effective when they are aligned with students' preferred ways of learning (Nasim & Mujeeba, 2021). This also aligns with research emphasizing learner autonomy and guided support in AI-mediated writing contexts (Estrella, 2025; Ghufroon & Rosyida, 2018).

Limitations

This study has several limitations. First, the sample was drawn from a single Saudi public university and included only preparatory-year male EFL students, which limits the generalizability of the findings to other institutions, proficiency levels, and learner groups. Second, the study relied on self-report questionnaire data, which reflect students' perceptions rather than their actual writing performance or actual Grammarly use. Third, the cross-sectional design does not permit causal interpretation. Finally, although the sample size was adequate for the analyses conducted, it may have limited the detection of smaller effects, particularly in the relationship analyses. These limitations

suggest that the findings should be interpreted as context-specific evidence about perceptions rather than as broad claims about the impact of Grammarly on EFL writing.

Future research could strengthen this line of inquiry by combining perception data with actual writing samples, revision behavior, or classroom observation. It would also be useful to include female learners, students from multiple institutions, and broader proficiency ranges to examine whether perceptions of Grammarly differ across learner groups. Mixed-methods or longitudinal designs could also help clarify how perceptions of usefulness, limitations, and overreliance evolve over time.

Conclusion

In conclusion, this study suggests that preparatory-year Saudi EFL students generally viewed Grammarly as a useful tool for correcting writing mechanics and improving writing efficiency. At the same time, their perceptions of the tool's limitations and their own overreliance were more mixed. The findings also showed that perceived usefulness was not significantly associated with perceived overreliance in this sample. Taken together, these results support the use of Grammarly as a pedagogical scaffold in EFL writing instruction, provided that its use is accompanied by guidance that encourages critical evaluation, balanced use, and continued development of learner autonomy.

Acknowledgments

This study is supported via funding from Prince Sattam bin Abdulaziz University project number PSAU/2026/R/1447.

References

- Abu Guba, M. N., Awad, A., & Abu Qub'a, A. (2024). Grammarly in teaching writing to EFL learners at low levels: How useful is it? *World Journal of English Language*, 14(3), 1–13. <https://doi.org/10.5430/wjel.v14n3p1>
- Al Mahmud, F. (2023). Investigating EFL students' writing skills through artificial intelligence: Wordtune application as a tool. *Journal of Language Teaching and Research*, 14(5), 1395–1404. <https://doi.org/10.17507/jltr.1405.28>
- Altamimi, D. H. F. (2025). Unlocking potential: Saudi EFL male students' perspectives on AI tools for enhancing English writing proficiency. *Arab World English Journal (AWEJ) Special Issue on Artificial Intelligence, No. 3*, 40-58. <https://doi.org/10.24093/awej/AI.3>
- Bahing, B., Meilantina, M., & Akhmad, W. (2025). Students' perceptions in using Grammarly and QuillBot to improve writing skills at MA Annur Teluk Palinget: Graduate students of Palangka Raya University. *Journal of Innovation Research and Knowledge*, 5(5), 5439–5450. <https://doi.org/10.53625/jirk.v5i5.11405>
- Calma, A., Cotronei-Baird, V., & Chia, A. (2022). Grammarly: An instructional intervention for writing enhancement in management education. *The International Journal of Management Education*, 20(3), 100704. <https://doi.org/10.1016/j.ijme.2022.100704>
- Dewi, U. (2023). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. *Lingua Cultura*, 16(2), 155–161. <https://doi.org/10.21512/lc.v16i2.8315>
- 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>
- 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. <https://doi.org/10.37256/ser.6120255896>
- Fitria, F. (2023). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. *Journal of English Language Teaching*, 14(2), 1–10. <https://doi.org/10.31002/metathesis.v5i1.3519>

- Fitria, R. A., & Sabarun, S. (2022). Students' perceptions of the use of Grammarly in undergraduate thesis writing. *Project: Professional Journal of English Education*, 5(2), 366–371. <https://doi.org/10.22460/project.v5i2.p366-371>
- Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a Foreign Language (EFL) writing. *Lingua Cultura*, 12(4), 395–403. <https://doi.org/10.21512/lc.v12i4.4582>
- 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). <https://doi.org/10.59652/jetm.v3i2.469>
- Huang, H.-W., Li, Z., & Taylor, L. (2020). The effectiveness of using Grammarly to improve students' writing skills. In *Proceedings of the 2020 the 5th International Conference on Distance Education and Learning (ICDEL 2020)*. <https://doi.org/10.1145/3402569.3402594>
- Ishtiaq, M., Shahid, S. H., Ishtiaq, R. M., & Nasim, S. M. (2025). Examining writing errors among Saudi EFL learners: Practices and perspectives. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan*, 44(2), 198–208. <https://doi.org/10.21831/cp.v44i2.70967>
- Khan, M. O., Nazim, M., & Alzubi, A. (2024). Exploring Arab EFL learners' attitudes: Is Grammarly a game-changer in academic writing classes? *Educational Administration: Theory and Practice*, 30(4), 1365–1371. <https://doi.org/10.53555/kuey.v30i4.1612>
- Lai, Z. C. C. (2025). The impact of AI-assisted blended learning on writing efficacy and resilience. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1), 1–21. <https://doi.org/10.4018/IJCALLT.377174>
- Martinez-Carrasco, R., & Chabert, A. (2023). Writing on steroids? Accuracy of automatic corrective feedback in L2 competence development. *Bellaterra Journal of Teaching & Learning Language & Literature*, 16(3), 1–25. <https://doi.org/10.5565/rev/jtl3.1142>
- Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency, and writing quality. *Discover Education*, 4, 170. <https://doi.org/10.1007/s44217-025-00602-7>
- Mohd Nasim, S. (2022). Metacognitive listening comprehension strategies of Arab English language learners. *Education Research International*, 2022, Article 9916727. <https://doi.org/10.1155/2022/9916727>
- Mohd Sham, N. H., Noordin, N., & Ahmad, N. K. (2025). Exploring students' perception towards the use of Grammarly to improve grammatical accuracy in essay writing. *Arab World English Journal*, 16(3), 227–243. <https://doi.org/10.24093/awej/vol16no3.13>
- Muhammad, G. (2024). The effectiveness of Grammarly features in building arguments in writing essays. *Jurnal Ilmu Komputer dan Teknologi (IKOMTI)*, 5(3), 41–47. <https://doi.org/10.35960/ikomti.v5i3.1664>
- Nasim, S. M., & Mujeeba, S. (2021). Learning styles of Saudi ESP students. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 13(4), 1–17. <https://doi.org/10.21659/rupkatha.v13n4.55>
- Nasim, S. M., & Mujeeba, S. (2024). Arab EFL students' and instructors' perceptions of errors in mechanics in second language paragraph writing. *FWU Journal of Social Sciences*, 18(1), 87–103. <https://doi.org/10.51709/19951272/Spring2024/7>
- Nasim, S. M., Mujeeba, S., AlShraah, S. M., Khan, I. A., Amir, & Ali, Z. (2024). Exploring pedagogical perspectives of EFL instructors: advantages, disadvantages, and implications of Blackboard as an LMS for language instruction. *Cogent Social Sciences*, 10(1), 2312659. <https://doi.org/10.1080/23311886.2024.2312659>
- Ozfidan, B., El-Dakhs, D. A. S., & Alsalm, L. A. (2024). The use of AI tools in English academic writing by Saudi undergraduates. *Contemporary Educational Technology*, 16(4), ep527. <https://doi.org/10.30935/cedtech/15013>
- Rafida, T., Suwandi, S., & Ananda, R. (2024). EFL students' perception in Indonesia and Taiwan on using artificial intelligence to enhance writing skills. *Jurnal Ilmiah Peuradeun*, 12(3), 987–1016. <https://doi.org/10.26811/peuradeun.v12i3.1520>
- Resiana, A. T., Zamzam, A., Putera, L. J., Amrullah, & Arfah, H. (2024). The effectiveness of Grammarly application on the students' argumentative writing progress. *Journal of English Education Forum*, 4(3), 153–159. <https://doi.org/10.29303/jeeef.v4i3.722>

- Tran, N. N. (2024). The use of Grammarly application in writing a graduation project among senior students majoring in English language at Nam Can Tho University. *International Journal of Science and Management Studies (IJSMS)*, 7(4), 23–30. <https://doi.org/10.51386/25815946/ijms-v7i4p103>
- Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, 15(3), 343. <https://doi.org/10.3390/educsci15030343>
- Wahyu, T., & Zur, S. (2023). EFL students' perceptions on Grammarly use in proofreading essay tasks. *AL LUGHAWIYAAT*, 4(2). <https://doi.org/10.31332/alg.v4i2.8046>
- Yang, H. (2018). Efficiency of online grammar checker in English writing performance and students' perceptions. *Korean Journal of English Language and Linguistics*, 18(3), 328–348. <https://doi.org/10.15738/kjell.18.3.201809.328>
- Yawiloeng, R. (2025). Investigating EFL students' writing problems through the lens of Grammarly. *Theory and Practice in Language Studies*, 15(7), 2172-2181. <https://doi.org/10.17507/tppls.1507.09>