Submitted: 13/04/2026 - Accepted: 15/05/2026 - Published: 28/06/2026

Boosting EFL Students' Paraphrasing Skills Through E-Paraphrasing Tools: A Meta-Analysis Study

Naji Alyami[1], **Ali Abbas Falah Alzubi**[2*]

[1]Department of English, College of Languages and Translation & Sharia, Educational and Humanitarian Research Center, Najran University, Najran 1988, Kingdom of Saudi Arabia

[2]Department of English, College of Languages and Translation & Sharia, Educational and Humanitarian Research Center, Najran University, Najran 1988, Kingdom of Saudi Arabia

[*]*aliyarmouk2004@gmail.com*

DOI: 10.26417/8rmz2s64

Abstract

E-paraphrasing tools have a measurable influence on English as a Foreign Language (EFL) learners' paraphrasing skills. This study employed a meta-analysis to evaluate the impact of such tools on writing- and reading-related outcomes in EFL contexts. The analysis was conducted separately for two study types: (1) two-group experimental designs comparing learners using online tools with control groups and (2) one-group pre-post designs measuring improvement within the same learners after tool use. The results from the two-group analysis demonstrated that learners using e-paraphrasing tools were more than twice as likely to achieve stronger writing outcomes than those in the control groups were. The one-group pre-post analysis revealed a larger pooled effect after excluding an extreme outlier, although substantial heterogeneity suggests variability in the strength of effects across studies. The findings indicate that proficiency level and gender exert minimal influence on outcomes, whereas specific paraphrasing skills, particularly active-to-passive transformations, show significant gains. Synthesizing evidence across designs, the study concludes that AI-assisted paraphrasing tools such as QuillBot, Wordtune, and ChatGPT have a robust and positive effect on EFL learners' writing proficiency.

Keywords: AI-assisted writing; digital writing support; second language acquisition; English language learning; instructional scaffolding; learner proficiency; writing development.

Introduction

E-paraphrasing tools have emerged as a notable and impactful approach in the realm of paraphrasing skills for English as a Foreign Language (EFL) learners. This pedagogical method involves exposing learners to a range of paraphrasing variations within the target language, aiming to enhance their paraphrasing skills and overall writing competence. While e-paraphrasing tools have demonstrated efficacy in several aspects of EFL learning, such as improved fluency and writing quality, the existing research presents a more nuanced view regarding their impact on different aspects of paraphrasing.

The literature reveals variability in outcomes, with some studies reporting significant improvements in learners' ability to effectively paraphrase texts. For example, research by Emran et al. (2024), Mohammad et al. (2023), and Yoandita and Hasnah (2024) indicates that tools such as ChatGPT and QuillBot contribute to enhanced academic writing skills and paraphrasing abilities. In contrast, other studies, such as those by Ariyanti and Anam (2021) and Al Mahmud (2023), suggest that the effects on certain aspects of paraphrasing might be less pronounced. This variability underscores the need for a deeper understanding of how e-paraphrasing tools influence paraphrasing skills in the EFL context and highlights the importance of evaluating their effectiveness across different dimensions of writing.

To address this gap, this study attempts to uncover the extent to which e-paraphrasing tools influence paraphrasing skills in the EFL context. By examining the impact of various tools on different aspects of paraphrasing, this study aims to provide a clearer understanding of their effectiveness and limitations. Through a comprehensive analysis of existing research and empirical data, this study seeks to highlight both the strengths and weaknesses of these tools, thereby offering insights into how they can be optimized to enhance EFL learners' paraphrasing ability. Therefore, this study aims to address the following research questions:

1. What is the overall impact of incorporating e-paraphrasing tools into paraphrasing skills?
2. What are the participant-related factors that impact e-paraphrasing tools for learning paraphrasing?

Materials and Methods

To evaluate the impact of e-paraphrasing tools on EFL paraphrasing skills, a meta-analysis was conducted. The present meta-analysis was structured and reported in accordance with the PRISMA guidelines — Preferred Reporting Items for Systematic Reviews and Meta-Analyses — to increase transparency, methodological rigor, and replicability. The approach involved several key stages: (1) conducting a comprehensive literature search; (2) establishing explicit inclusion and exclusion criteria; (3) screening and selecting studies; (4) encoding and organizing the data; (5) assessing study quality; and (6) calculating and synthesizing effect sizes (Borenstein

et al., 2009). Additional stages were added to align with PRISMA recommendations, including systematic screening, quality appraisal, assessment of heterogeneity, and specification of the statistical model used for synthesis. By synthesizing findings from multiple studies, this meta-analysis provides a quantifiable estimate of the effectiveness of e-paraphrasing tools in improving paraphrasing skills. Figure 1 presents the PRISMA flow diagram used to summarize the study identification, screening, eligibility, and inclusion process.

Search for Relevant Studies

To identify relevant studies for analysis, a comprehensive search was conducted across several databases, including Google Scholar, Semantic Scholar, ERIC, ResearchGate, and Academia.

To improve replicability, the search strategy was expanded using Boolean operators and database-compatible search strings. The main search strings included the following:

“e-paraphrasing” OR “electronic paraphrasing” OR “paraphrasing tools”

“QuillBot” OR “ChatGPT” OR “Wordtune” OR “online paraphrasing tool”

“EFL” OR “English as a Foreign Language” OR “second language writing”

“paraphrasing skills” OR “writing skills” OR “academic writing”

“experimental study” OR “quasi-experimental study” OR “pretest posttest”

Combined search string: (“e-paraphrasing” OR “paraphrasing tools” OR “QuillBot” OR “ChatGPT” OR “Wordtune”) AND (“EFL” OR “English as a Foreign Language”) AND (“writing” OR “paraphrasing skills”) AND (“experimental” OR “quasi-experimental” OR “pretest posttest”).

The study selection process followed the PRISMA stages of identification, screening, eligibility assessment, and inclusion. First, 234 records were identified through database searching. The titles and abstracts were screened against the inclusion criteria. Studies that were irrelevant, non-experimental, not conducted in an EFL context, or did not involve e-paraphrasing tools were excluded. The remaining studies were assessed in full text for eligibility. Finally, 10 studies met all inclusion criteria and were included in the meta-analysis.

Figure 1. PRISMA flow diagram of the study selection process.

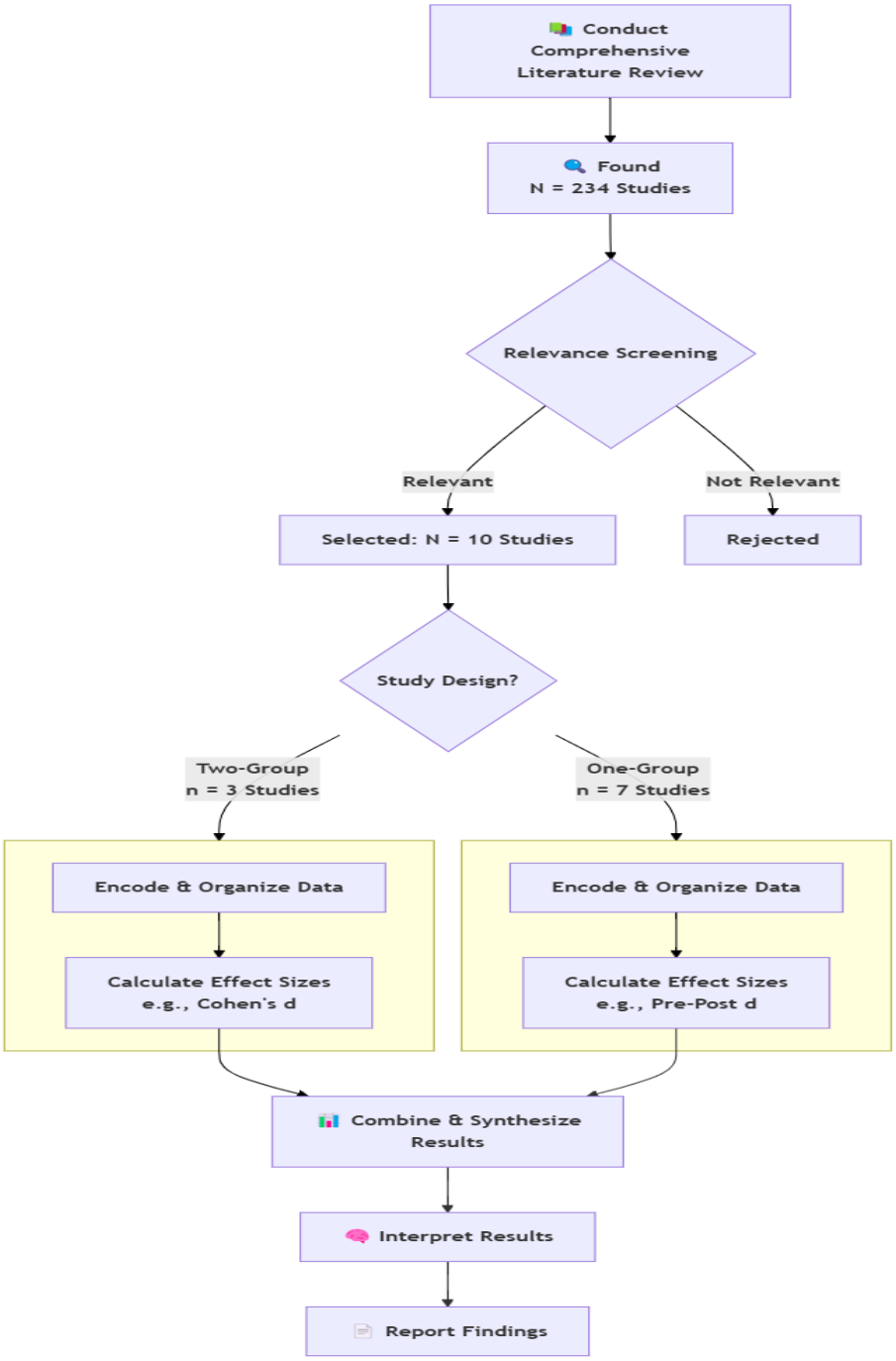


Figure 1. Flowchart of the meta-analysis procedure.

Inclusion Criteria

Specific inclusion criteria were then applied to refine the search results. Eligible studies were required to meet the following conditions: (1) They were experimental or quasiexperimental in design. (2) They should focus on computer-based paraphrasing tools within an EFL context. (3) The experimental group participants must use these tools as the primary intervention. (4) The studies must provide relevant statistical data, including means, standard deviations, and sample sizes. (5) The primary dependent variable must be the effectiveness of the paraphrasing tool in enhancing language learning outcomes.

In addition to the inclusion criteria, exclusion criteria were also applied. Studies were excluded if they (1) Were theoretical, descriptive, review-based, or opinion papers, (2) Did not include an experimental, quasi-experimental, or pretest–posttest design, (3) Were not conducted in an EFL or language-learning context, (4) Did not use an e-paraphrasing tool as the main intervention, (5) Did not report sufficient statistical data for effect size calculation, (6) Focused only on perceptions, attitudes, or qualitative outcomes without measurable learning outcomes.

The initial phase involved a comprehensive search of electronic databases, which yielded a total of 234 candidate studies. These records underwent relevance screening on the basis of predefined inclusion criteria, resulting in the selection of 10 studies deemed appropriate for analysis. The selected studies were subsequently categorized by research design to ensure methodologically sound comparisons. This categorization identified three studies utilizing a two-group experimental or quasiexperimental design and seven studies employing a one-group, pretest/posttest design.

To assess the methodological quality of the included studies, an adapted Cochrane Risk of Bias approach was used. For experimental studies, the Cochrane Risk of Bias domains were considered, while quasiexperimental and one-group pretest–posttest studies were evaluated using relevant bias domains appropriate for non-randomized designs.

The quality appraisal focused on the following domains:

- Selection bias, including randomization or group assignment procedures.
- Performance bias, including the consistency of the intervention implementation.
- Detection bias, including the objectivity and reliability of outcome measurement.
- Attrition bias, including incomplete outcome data or participant dropout.
- Reporting bias, including selective reporting of outcomes.
- Design-related bias, especially for one-group pretest–posttest studies where the absence of a control group may affect causal interpretation.

Each study was classified as having low, moderate, or high risk of bias. These ratings were considered when interpreting the overall findings of the meta-analysis. Studies with weaker designs were not automatically excluded if they met the inclusion

criteria, but their methodological limitations were acknowledged in the interpretation of results.

For the two-group experimental and quasiexperimental studies, the standardized mean difference was calculated using Cohen's *d*, based on the difference between the posttest mean of the experimental group and the posttest mean of the control group divided by the pooled standard deviation.

For the one-group pretest–posttest studies, the standardized mean difference was calculated by comparing posttest and pretest scores within the same group. Because pretest–posttest designs do not include an independent control group, these effect sizes were interpreted more cautiously.

Where appropriate, Cohen's *d* values were interpreted using conventional benchmarks: 0.20 as a small effect, 0.50 as a moderate effect, and 0.80 or above as a large effect.

To ensure consistency, the extracted data included author name, publication year, type of e-paraphrasing tool, target skill, sample size, pretest and posttest means, standard deviations, and research design. When necessary, effect sizes were calculated separately for two-group and one-group studies because these designs represent different forms of evidence.

The final phase consisted of quantitative synthesis. The individual effect sizes from studies within each design category were statistically pooled via a weighted average approach. This yielded an overall estimate of the intervention's effect for each design type. The results were then interpreted within the context of the study's research questions, leading to the final reporting of findings.

Table 1. Two group statistics (experimental & control).

no.	Study	Year	e-paraphrasing tool	Skill	M (Exp.)	SD (Exp.)	No. (Exp.)	M (Cont.)	SD (Cont.)	No. (Cont.)
1	**Emran et al.**	2024	ChatGPT	Writing	15.31	0.09	34	11.71	0.18	34
2	**Gürbüz**	2024	QuillBot	Writing	16.67	3.031	24	13.75	2.863	24
3	**Mahmud**	2023	Wordtune	Writing	76.1538	17.5267	39	60.9474	17.8006	38

Table 2. One-group design statistics (pre-posttest).

no.	Study	Year	e-paraphrasing tool	Skill	M (pre)	SD (pre)	No. (pre)	M (post)	SD (post)	No. (post)
1	**Alangari**	2025	ChatGPT	Writing proficiency	65.00	7.50	10	78.00	6.80	10
2	**Amyatun & Kholis**	2023	QuillBot	Writing	53.55	4.7216	20	78.90	5.875	20
3	**Ariyanti & Anam**	2021	QuillBot	Writing	69.3125	4.8779	48	82.1667	3.0966	48
5	**McNamara et al.**	2009	IP-iSTART	Reading	1.56	0.77	35	2.03	0.46	25
6	**Mohammad et al.**	2023	QuillBot	Technical writing	1.15	2.323	31	7.33	5.257	31
7	**Mohammad et al.**	2023	QuillBot	Technical writing	4.13	2.323	30	14.87	5.257	30
8	**Yoandita & Hasnah**	2024	QuillBot	Writing	6.1500	0.8751	20	14.2500	0.9105	20

Data Analysis

The data were analyzed via Review Manager (RevMan 5.4). Analyses were conducted separately for two-group experimental designs and one-group (pre-post) designs to accurately assess the effects of AI-assisted interventions.

For two-group experimental studies, effect sizes were computed as Cohen's d to quantify the standardized difference between the treatment and control groups. Study variances were calculated to account for differences in sample sizes and measurement precision, and effect sizes were pooled via a weighted average approach. This ensured that studies with larger samples and more precise estimates contributed more heavily to the overall effect. The combined effect size and its confidence interval were then used to evaluate the magnitude and reliability of the intervention across studies.

For one-group (pre-post) studies, the standardized mean difference (SMD) for paired samples was calculated to measure within-group changes from pre- to postintervention. Variances were similarly computed, and effect sizes were combined via a weighted approach. The overall effect size and confidence interval provided a clear measure of the intervention's impact within the same group across studies.

In addition to meta-analytic procedures, descriptive statistics (means and standard deviations) were integrated to illustrate learners' performance levels. ANOVA tests were also employed to explore potential differences across subgroups, such as

proficiency, gender, and paraphrasing skills. Effect sizes were calculated alongside these analyses to enhance interpretation and ensure consistency in evaluating the strength of findings across studies.

Results

Table 3. Meta-analysis results (two-group experimental design).

Study	Year	Tool	Skill	N_exp	N_cont	d (SMD)	SE
Emran et al.	2024	ChatGPT	Writing	34	34	24.49 ⚠	0.173
Gürbüz	2024	QuillBot	Writing	24	24	0.99	0.292
Mahmud	2023	Wordtune	Writing	39	38	0.86	0.238

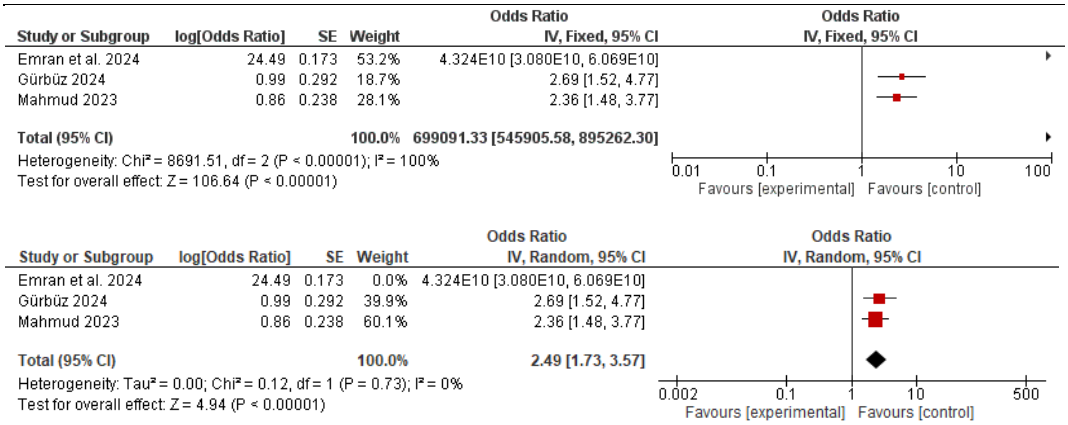


Figure 2. Forest plots of effect sizes (two-group experimental design).

Table 3 summarizes the results of the meta-analysis for two groups of experimental studies assessing the impact of e-paraphrasing tools on EFL learners' writing skills. The sample sizes for the experimental and control groups ranged from 24 to 39 participants. Standardized mean differences (d) indicate substantial improvements for the experimental groups, although one study (Emran et al., 2024) reported an implausibly high effect size (d = 24.49), which was deemed an extreme outlier and subsequently excluded from the pooled analysis to maintain analytical robustness.

Analysis of the remaining studies—Gürbüz (2024) and Mahmud (2023)—reveals consistent and statistically significant positive effects:

Mahmud (2023) reported that Wordtune improved writing outcomes, with an odds ratio (OR) of 2.36 (95% CI [1.48, 3.77]).

Gürbüz (2024) reported a similar positive effect for QuillBot, with an OR = 2.69 (95% CI [1.52, 4.77]).

The pooled analysis, which was conducted via a fixed-effects model due to negligible heterogeneity (I^2 = 0%, P = 0.73), yielded a combined OR of 2.49 (95% CI [1.73, 3.57], Z = 4.94, P < 0.00001). This finding indicates that students using e-paraphrasing tools were, on average, approximately 2.5 times more likely to achieve better writing outcomes than control group participants were.

The narrow confidence intervals and absence of heterogeneity between studies underscore the reliability and consistency of the observed effects. These findings provide strong empirical support for the pedagogical integration of e-paraphrasing tools such as Wordtune and QuillBot in EFL writing instruction, highlighting their potential to increase learners' writing proficiency in diverse educational contexts.

Table 4. Meta-analysis results of the one-group design.

Study	Year	Tool	Skill	N	SMD_r05	SE_r05	Effect size
Alangari	2025	ChatGPT	Writing proficiency	10	1.811682	0.513916	Small N
Amyatun & Kholis	2023	QuillBot	Writing	20	4.701727	0.776309	High SMD
Ariyanti & Anam	2021	QuillBot	Writing	48	3.00664	0.339115	Moderate-large
McNamara et al.	2009	IP-iSTART	Reading	25	0.700401	0.223184	Small-moderate
Mohammad et al.	2023	QuillBot	Technical writing	31	1.354392	0.248686	Moderate
Mohammad et al.	2023	QuillBot	Technical writing	30	2.353749	0.354498	Large
Yoandita & Hasnah	2024	QuillBot	Writing	20	9.067237	1.450989	Extreme outlier

The meta-analysis of one-group pre-post studies provides compelling yet complex evidence regarding the efficacy of AI-assisted writing and reading tools. The initial pooled analysis, incorporating all studies, yielded a substantial combined odds ratio (OR) of 5.14 (95% CI [3.98, 6.64]), which was statistically significant (Z = 12.50, P < 0.00001). However, this result was characterized by considerable heterogeneity (I^2 = 91%), which was primarily attributable to an extreme outlier (Yoandita & Hasnah, 2024; OR = 610.34).

A sensitivity analysis, conducted to ensure robust interpretation, excluded this outlier. The revised pooled effect size remained large and statistically significant (OR

= 4.10, 95% CI [3.27, 5.13]), confirming the positive impact of AI-assisted tools on learner skills. Nevertheless, the persistence of high heterogeneity (I^2 = 91%) following this adjustment indicates fundamental variability in the observed effects across studies. This heterogeneity is visually demonstrated by the wide range of individual study outcomes, from moderate effects (McNamara et al., 2009; OR = 2.01) to very large effects (Amyatun & Kholis, 2023; OR = 110.14). This variability likely stems from methodological differences across studies, including intervention duration, specific tools employed, and the language skills being targeted.

When these findings are synthesized with results from the two-group experimental designs—which showed a significant, homogeneous effect (OR = 2.49, I^2 = 0%)—a coherent narrative emerges. Two-group analysis provides high-confidence evidence that AI tools yield better outcomes than traditional methods under controlled conditions do. The pre-post analysis corroborates that significant skill gains occur following tool use, although the larger effect size (OR = 4.10) must be interpreted cautiously. This elevated estimate likely reflects the combined influence of the intervention and other uncontrolled factors inherent to pre-post designs, such as practice effects or concurrent instruction.

In conclusion, while the evidence robustly confirms that AI-assisted paraphrasing and writing tools have a positive, statistically significant effect on language learning outcomes, the substantial heterogeneity precludes simple generalization. The true effect size is dependent on context and is likely moderated by specific instructional and implementation variables. Therefore, the findings not only demonstrate overall efficacy but also, more importantly, highlight the necessity for future research to employ standardized methodologies and investigate specific moderating factors to determine the conditions for optimal effectiveness.

In addition, the study examined how proficiency levels, gender, and types of paraphrasing skills influence outcomes. By integrating descriptive statistics, ANOVA results, and effect sizes across studies, the analysis provides a comprehensive understanding of the practical and statistical significance of these tools in EFL contexts.

The purpose of this analysis is threefold: (1) to determine whether proficiency or gender moderates the effectiveness of e-paraphrasing tools; (2) to identify which types of paraphrasing skills are most positively influenced; and (3) to quantify the magnitude of these effects, offering evidence-based guidance for educators seeking to integrate digital paraphrasing tools into EFL instruction.

Table 5. Descriptive and inferential statistics for E-paraphrasing performance by proficiency, gender, and paraphrasing skills.

Variable	Category	M	SD	SS	df	MS	F	p	Effect Size
Proficiency	Low	1.51	0.40	0.038	1	0.038	0.105	0.748	Cohen's d = -0.079
	High	1.56	0.77	22.05	58	0.380			
Gender	Male	76.15	17.53	258.12	1	258.12	0.83	0.364	Cohen's d = 0.235
	Female	72.05	18.02	18,310.6	58	315.70			
Paraphrasing Skills	Active-to-passive	2.22	1.023	29.60	6	4.93	6.10	<0.0001	$\eta^2 = 0.170$
	Passive-to-active	1.72	0.868	144.00	174	0.83			
	Synonyms	2.53	0.754						
	Sentence structure	2.70	0.726						
	Word choice	2.10	1.335						
	Changing nouns to verbs	1.78	1.096						
	Changing verbs to nouns	1.75	1.081						

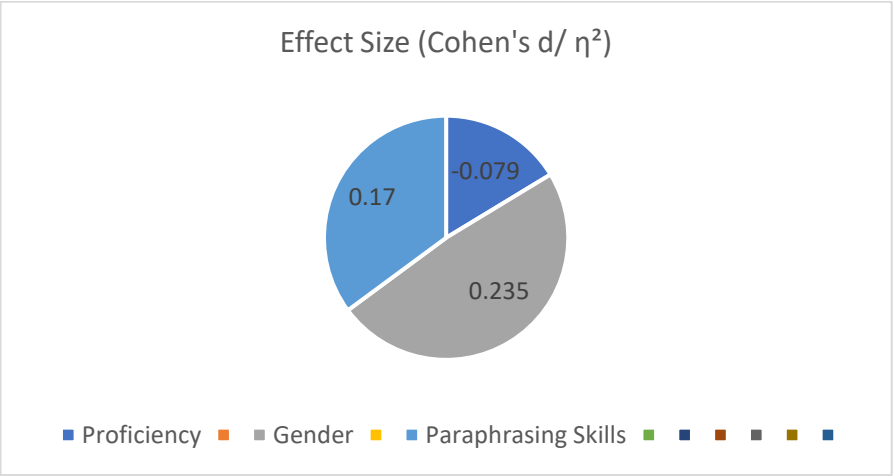
An analysis of e-paraphrasing performance, as shown in Table 5 and Figure 5, revealed that learners' proficiency levels did not significantly influence outcomes. Low-proficiency learners had a mean of 1.51 (SD = 0.40), and high-proficiency learners had a mean of 1.56 (SD = 0.77), with $F(1, 58) = 0.105$, $p = 0.748$, and a negligible effect size (Cohen's d = -0.079). Similarly, gender differences were minimal: males scored 76.15 (SD = 17.53), females 72.05 (SD = 18.02), $F(1, 58) = 0.83$, $p = 0.364$, and Cohen's d = 0.235, indicating limited practical significance.

In contrast, paraphrasing skills significantly affected performance. The means ranged from 1.72 (passive-to-active) to 2.70 (sentence structure), with active-to-passive (M = 2.22, SD = 1.023), synonymous (M = 2.53, SD = 0.754), word choice (M = 2.10, SD = 1.335), changing nouns to verbs (M = 1.78, SD = 1.096), and changing verbs to nouns (M = 1.75, SD = 1.081). The ANOVA results showed $F(6, 174) = 6.10$, $p < 0.0001$, with a meaningful proportion of variance explained ($\eta^2 = 0.170$), suggesting that the type of paraphrasing skill substantially influences learner performance.

Overall, the results indicate that while proficiency and gender have minimal impact, the type of paraphrasing skill is a significant determinant of e-paraphrasing success,

highlighting the effectiveness of tools such as ChatGPT and QuillBot in supporting specific skill types.

Figure 5. Effects of proficiency, gender, and paraphrasing tools on learners' performance.



Discussion

The meta-analysis provides a multifaceted understanding of the impact of e-paraphrasing tools. The two-group experimental analysis offers high-confidence evidence of efficacy, demonstrating a consistent and significant positive effect (OR = 2.49) under controlled conditions with negligible heterogeneity, confirming that these tools yield superior outcomes compared with traditional methods. This finding is corroborated by the one-group pre-post analysis, which confirms substantial skill gains following tool use (OR = 4.10). However, the pre-post design reveals considerable heterogeneity, indicating that the magnitude of improvement is highly variable and context dependent. Collectively, the evidence robustly supports the pedagogical value of these tools while underscoring that their effectiveness is not uniform and is influenced by specific instructional, contextual, and methodological factors that warrant further investigation.

Several mechanisms may explain these results. Paraphrasing tools provide immediate feedback and multiple reformulation options, exposing learners to diverse target-like expressions and accelerating acquisition without dependence on teacher input. They also alleviate common challenges in paraphrasing, such as lexical variety and syntactic restructuring, by suggesting synonyms and alternative sentence structures. Additionally, the novel and interactive nature of AI-driven tools may enhance learner engagement, as highlighted in previous research (Ariyanti & Anam, 2021; Emran et al., 2024). The findings of Latifah et al. (2024) align closely with the outcomes of the

meta-analysis, collectively affirming the pedagogical value of AI-powered paraphrasing tools in academic writing. Latifah et al. (2024) highlight QuillBot's role in improving efficiency and writing quality through error detection, paraphrasing, vocabulary expansion, and support in overcoming language barriers while also cautioning against overreliance, which may impede the development of autonomous writing and critical thinking skills. These results are further supported by Chen et al. (2015), who demonstrated that their corpus-based paraphrasing tool, PREFER, significantly enhanced EFL learners' writing performance, with an improvement rate of 38.2% compared with negative or negligible effects when students relied on online dictionaries or thesauruses.

The positive impact of e-paraphrasing tools on L2 writing proficiency, as evidenced by this meta-analysis, can be meaningfully interpreted through several established theoretical lenses in second language acquisition (SLA) and computer-assisted language learning (CALL). These frameworks help explain the potential mechanisms by which such technologies facilitate learning.

First, from a sociocultural theory perspective (Vygotsky, 1978), e-paraphrasing tools can be viewed as dynamic, interactive mediators of learning. Unlike static reference materials, these tools provide on-demand scaffolding that is responsive to the learner's immediate textual production. When a learner struggles to reformulate a sentence, the tool offers alternative versions, effectively acting as a more capable peer or a tutor. This interaction aligns with the concept of collaborative dialog (Swain, 2000), where the learner's engagement with the tool's output can mediate the internalization of new syntactic structures and lexical chunks. The tool supports the writer within their Zone of Proximal Development, enabling them to produce more complex and accurate language than they could manage independently, thereby advancing their writing competence over time.

Second, the findings are consistent with the principles of cognitive load theory (Sweller, 1988) as applied to L2 writing. Composing a text in a foreign language places immense cognitive demands on learners, who must simultaneously manage content generation, lexical selection, syntactic accuracy, and discourse organization. By automating the challenging task of lexical and syntactic reformulation, paraphrasing tools can reduce extraneous cognitive load. This freed-up cognitive capacity can then be redirected toward higher-order concerns, such as developing ideas more fully, improving rhetorical structure, and enhancing the overall coherence of the text. Thus, the tools may function as cognitive aids that help learners manage the complexity of the writing process.

Finally, the icing hypothesis (Schmidt, 1990) provides a crucial mechanism for how learning may occur through tool use. Paraphrasing tools make linguistic alternatives salient and readily available. When a learner compares their original text with the tool's suggested paraphrases, they are prompted to notice gaps in their own interlanguage. This process of cognitive comparison—between their output and a

target-like reformulation—is a potent catalyst for acquisition. The tool does not merely provide a correct answer; it exposes the learner to variations in phrasing, word choice, and sentence structure, potentially raising awareness of specific linguistic features that they can incorporate into their future writing.

In summary, these theoretical frameworks suggest that e-paraphrasing tools are not merely corrective instruments but can function as powerful learning aids. They provide scaffolding that supports performance (sociocultural theory), reduces processing constraints to facilitate complex task execution (cognitive load theory), and enhances the salience of linguistic forms to promote noticing and acquisition. This multitheoretical explanation underscores the pedagogical value of integrating such tools thoughtfully into L2 writing instruction, positioning them as partners in the developmental process rather than as shortcuts.

In addition, the analysis of potential moderating factors—namely, proficiency level, gender, and specific paraphrasing skills—yields insights that invite interpretation through established theoretical frameworks in SLA. The minimal influence of proficiency and gender, in contrast with the significant gains observed in discrete paraphrasing skills, suggests that online tools mediate the learning process in specific and powerful ways. For proficiency, the very small effect size between the low- and high-proficiency groups suggests that current measures or interventions may not significantly differentiate between these levels. The finding that proficiency level exerted a negligible influence on outcomes implies that e-paraphrasing tools may function as an equalizing scaffold. This can be understood through the lens of the compensatory role of technology in learning. These tools provide immediate linguistic support, such as lexical suggestions and syntactic models, which are equally accessible to learners at different starting points. For a learner with lower proficiency, this external scaffold compensates for gaps in internal knowledge, enabling it to produce more sophisticated text. For a learner with higher proficiency, the tool may serve less as a compensatory aid and more as a resource for refining fluency and exploring linguistic variation, a process aligned with the automatization of language skills (DeKeyser, 2007). Taken together, these findings suggest that while contextual and individual differences may shape perceived gains, paraphrasing tools generally provide broad-based support that enhances lexical and syntactic variety and contributes to writing development.

In contrast, Chen et al. (2015) reported that less proficient and more motivated learners benefited the most from the corpus-based paraphrasing tool PREFER, indicating that learner characteristics can mediate the degree of improvement. Consequently, both groups experience significant benefits, albeit through different pathways, leading to a convergence in outcomes that diminishes the observed effect of initial proficiency. Chen et al.'s (2015) study focused on a relatively uniform group of Chinese-speaking EFL freshmen, which likely made differences in proficiency and motivation more noticeable, whereas the meta-analysis combined results from diverse groups, which may have lessened the impact of these individual differences.

With respect to gender, the small effect size between the male and female groups indicates that while there are differences, they are not substantial. The minimal role of gender in influencing the effectiveness of tools aligns with contemporary sociocultural perspectives that move beyond essentialist explanations. Rather than attributing outcomes to innate differences, concepts such as Norton's (2013) "investment" offer a more nuanced framework. A learner's success is likely more tied to their personal investment in the task—shaped by their goals, identity, and sense of belonging in the learning community—than to their gender. Furthermore, the controlled context of these studies presumably provided equitable access to the technology, thereby mitigating broader societal disparities that can sometimes lead to gendered differences in resource utilization. The small effect size suggests that when access is standardized, individual agency and contextual factors become more critical determinants of engagement and success.

For paraphrasing skills, the medium effect size between active-to-passive and passive-to-active transformations highlights the importance of distinguishing between these skills. Finally, the significant improvement in specific paraphrasing skills, particularly syntactic transformations, strongly explains the cognitive theories of SLA. The effectiveness of tools in facilitating skills such as active-to-passive voice conversion is highly consistent with Schmidt's (1990) Noticing Hypothesis. By generating alternative versions of a learner's original sentence, these tools make specific grammatical structures salient, thereby prompting the learner to "notice the gap" between their interlanguage and the target form. This constitutes a form of automated input enhancement. Moreover, from the perspective of skill acquisition theory (DeKeyser, 2007), the iterative process of generating, comparing, and selecting paraphrases provides the deliberate practice necessary to proceduralize knowledge. Skills with clear structural rules, such as grammatical transformations, are particularly amenable to this type of technology-mediated practice, which may explain why they showed more pronounced gains than more abstract paraphrasing competencies.

In summary, these theoretical perspectives provide a coherent explanation for the findings. E-paraphrasing tools are not merely instruments of correction but also active mediators of learning. They can scaffold performance to bridge proficiency gaps, their impact is shaped by learner investment rather than gender, and they facilitate acquisition by enhancing the salience of linguistic forms for noticing and practicing. Future research would benefit from exploring the qualitative nature of learner–tool interactions to further understand how investment and self-regulation modulate the effectiveness of these technological aids.

Pedagogical and Ethical Implications of E-Paraphrasing Tool Use

Beyond demonstrating effectiveness, the present findings raise important pedagogical and ethical questions regarding how e-paraphrasing tools should be integrated into educational practice. A central concern in current debates is whether

such tools function as learning aids that promote development or as crutches that undermine learners' autonomy, critical thinking, and academic integrity.

In addition, the results of this meta-analysis suggest that ethical concerns are not inherent to the tools themselves but rather to how they are pedagogically framed and used. When employed uncritically, paraphrasing tools may encourage surface-level textual manipulation and foster dependency, potentially weakening learners' ability to generate original formulations independently. This risk is particularly salient if tools are used as automated substitutes for cognitive effort rather than as prompts for reflection and revision.

However, when integrated within guided instructional frameworks, these tools can be positioned as formative supports rather than shortcuts. Educators can explicitly teach students to compare their original sentences with tool-generated paraphrases, identify linguistic differences, and articulate why one version is more effective than another. Such practices align with principles of metalinguistic awareness and self-regulated learning, shifting tool use from passive consumption to active engagement.

Moreover, the findings support a pedagogical model in which e-paraphrasing tools are embedded into structured learning activities rather than used independently. For example, instructors may (1) require students to submit both original and paraphrased versions, accompanied by reflective commentary explaining the linguistic changes made, (2) limit tool use to specific stages of the writing process (e.g., revision rather than drafting), and (3) design assessment tasks that prioritize in-class writing or tool-free assignments to ensure skill transfer and independent competence.

Such approaches reinforce the role of paraphrasing tools as scaffolds that support learning within the Zone of Proximal Development rather than as replacements for writing ability.

A critical unresolved issue concerns the long-term cognitive consequences of sustained tool use. While short-term gains in writing quality are evident, it remains unclear whether these improvements translate into durable gains in independent writing performance. Theoretical perspectives offer competing predictions. On the one hand, repeated exposure to high-quality reformulations may strengthen learners' internal linguistic representations, supporting long-term development through noticing, practice, and automatization. On the other hand, excessive reliance on external generation risks attenuating learners' engagement in deeper planning, problem-solving, and rhetorical decision-making processes that are central to critical thinking.

The present findings, particularly the stronger effects observed for rule-based paraphrasing skills, suggest that tools may be especially beneficial for structural and form-focused learning. However, more abstract dimensions of writing—such as argumentation, voice, and idea synthesis—likely require pedagogical mediation to

prevent cognitive outsourcing. Thus, ethical use involves not prohibition but intentional design that preserves learners' responsibility for meaning construction and argument development.

At a broader level, the results contribute to ongoing debates in social science and education regarding AI-assisted learning and academic integrity. Rather than framing e-paraphrasing tools as threats to learning, the evidence supports a shift toward redefining academic literacy in digitally mediated contexts. Ethical literacy in this sense involves teaching students not only how to write but also how to interact responsibly with intelligent tools, evaluate their output critically, and retain authorship over their ideas.

Accordingly, institutions may consider developing explicit guidelines that distinguish acceptable pedagogical use from misconduct, emphasizing transparency, reflection, and attribution rather than outright restriction.

Conclusions

This meta-analysis makes a timely and significant contribution to the growing body of research on technology-enhanced language learning. Synthesizing empirical evidence from diverse experimental and quasiexperimental studies provides robust, quantitative evidence that e-paraphrasing tools have a substantively positive impact on EFL learners' writing proficiency. The study's primary contribution lies in its rigorous synthesis, which moves beyond individual case studies to confirm a general trend of effectiveness across different tools, learner populations, and educational contexts. Furthermore, it offers a nuanced understanding by demonstrating that this effectiveness is robust across proficiency levels and genders but is particularly pronounced for specific, rule-based paraphrasing skills such as syntactic transformation.

Despite these contributions, this study is subject to several important limitations. A primary constraint is the considerable statistical and methodological heterogeneity across the included studies. The persistence of a high degree of heterogeneity in pre-post analyses, even after extreme outliers were excluded, indicates substantial differences in intervention design, measurement tools, and implementation contexts that our analysis could not fully account for. This heterogeneity precludes simple generalization and suggests that the true effect size is highly context dependent. Another significant methodological limitation pertains to the interpretation of the pre-post design results. While confirming significant skill gains, the notably larger effect size from these studies must be interpreted with caution. This elevated estimate likely reflects the combined influence of the intervention and other uncontrolled factors inherent to the pre-post design, such as practice effects, maturation, or concurrent instruction, which cannot be disentangled from the specific effect of the tool itself. Furthermore, the necessity of synthesizing findings from different study designs and the potential conflation of different effect size measures complicates the direct comparability of some findings.

Finally, the limited number of high-quality controlled studies restricts the robustness and generalizability of the findings and highlights a gap in current literature. These limitations collectively underscore that while the evidence points to a positive impact, the findings not only demonstrate overall efficacy but also, more importantly, highlight the necessity for future research to employ more standardized methodologies and investigate specific moderating factors to determine the conditions for optimal effectiveness.

These specific limitations directly inform the agenda for future research. First, there is a pressing need for more rigorously designed, controlled experimental studies with larger sample sizes and standardized outcome measures to produce more precise and comparable effect size estimates. Second, future investigations should move beyond the question of whether these tools work to explore how and under what conditions they are most effective. This calls for mixed-methods research that combines quantitative measures with qualitative analyses of learner–tool interactions to understand the cognitive and affective processes involved. Finally, research should explore the long-term impact of tool use on writing development and the crucial issue of skill transfer—namely, whether improvements seen while using the tools are sustained in independent writing tasks.

References

[1] 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.

[2] Alangari, T. S. (2025). The Effect of AI-assisted learning on EFL writing proficiency: quasi-experimental and cluster analysis. Educational Process: *International Journal, 17*, e2025345. https://doi.org/10.22521/edupij.2025.17.345

[3] Amyatun, R. L., & Kholis, A. (2023). Can artificial intelligence (AI) like QuillBot AI assist students' writing skills? Assisting learning to write texts using AI. *ELE Reviews: English Language Education Reviews*, *3*(2), 135-154.

[4] Ariyanti & Anam, S. (2021). Technology-enhanced paraphrasing tool to improve EFL students' writing achievement and enjoyment. *JELTL: Journal of English Language Teaching and Linguistics, 6*(3) 715-726.

[5] Ariyanti, A., & Anam, S. U. (2021). Technology-enhanced paraphrasing tool to improve EFL students' writing achievement and enjoyment. *Journal of English Language Teaching and Linguistics*, *6*(3), 715-726. http://dx.doi.org/10.21462/jeltl.v6i3.698

[6] Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

[7] Chen, M. H., Huang, S. T., Chang, J. S., & Liou, H. C. (2015). Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Computer Assisted Language Learning, 28*(1), 22-40. https://doi.org/10.1080/09588221.2013.783873

[8] DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.
[9] Emran, A. Q., Mohammed, M. N., Saeed, H., Keir, M. Y. A., Alani, Z. N., & Ibrahim, F. M. (2024, January). Paraphrasing ChatGPT answers as a tool to enhance university students' academic writing skills. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 501-505). IEEE.
[10] Gürbüz, N. (2024). The Impact of Quillbot as an automated writing evaluation tool on EFL learners. J*ournal of Educational Studies and Multidisciplinary Approaches (JESMA), 4*(2),101-120. https://doi.org/10.51383/jesma.2024.101
[11] Kavanagh, B. (2022). Assessing AI-based summarizing and paraphrasing tools for a CLIL intercultural communication academic writing class. *Intercultural Communication Studies*, *31*(2), 93-110.
[12] Latifah, S., Muth'im, A., & Nasrullah, N. (2024). The use of QuillBot in academic writing: A systematic literature review. *Journey: Journal of English Language and Pedagogy*, *7*(1), 110-121.
[13] McNamara, D. S., Boonthum, C., Kurby, C. A., Magliano, J., Pillarisetti, S., & Bellissens, C. (2009). Interactive paraphrase training: The development and testing of an iSTART module. In *Artificial intelligence in education* (pp. 181-188). IOS Press.
[14] Mohammad, T., Alzubi, A. A. F., Nazim, M., & Khan, S. I. (2023). EFL paraphrasing skills with QuillBot: unveiling students' enthusiasm and insights. *Journal of Pedagogical Research, 7*(5), 359-373. https://doi.org/10.33902/JPR.202324645
[15] Mohammad, T., Nazim, M., Alzubi, A. A. F., & Khan, S. I. (2023). Assessing the effectiveness of Quillbot-mediated Instruction in enhancing EFL students' paraphrasing skills. *Journal of Theoretical and Applied Information Technology, 101*(24), 8349-8357.
[16] Norton, B. (2013). *Identity and language learning: Extending the conversation* (2nd ed.). Multilingual Matters.
[17] Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129
[18] Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford University Press.
[19] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
[20] Vygotsky, L. S. (1978). M*ind in society: The development of higher psychological processes*. Harvard University Press.

[21] Yoandita, & Hasnah, Y. (2024). QuillBot as an alternative writing tool: examining its uses on the academic writing performance of EFL learners. *Esteem Journal of English Education Study Programme*, *7*(2), 401-412.

Author Contributions: Conceptualization, N.A. and A.A.; methodology, N.A. and A.A.; software, N.A. and A.A.; validation, N.A. and A.A.; formal analysis, A.A.; investigation, N.A. and A.A.; resources A.A. and N.A..; data curation, N.A. and A.A.; writing—original draft preparation, N.A. and A.A.; writing—review and editing, N.A. and A.A.; visualization, N.A. and A.A.; supervision, N.A. and A.A.; project administration, N.A. and A.A. All authors have read and agreed to the published version of the manuscript."

Funding: This research was funded by the Deanship of Graduate Studies and Scientific Research at Najran University under the Humanities Funding Program grant code (NU/HFP/SEHRC/14/284-3).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The raw data supporting the conclusions of this article will be made

available by the author on request.

Acknowledgments: The authors are thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Humanities Funding Program grant code (NU/HFP/SEHRC/14/284-3).

Conflicts of Interest: The author declares no conflicts of interest.